PTO 1542
(4 85)

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE

ONLINE SEARCH REQUEST FORM

1-120

USER J. Chambers SERIAL NUMBER 08/319,745

ART UNIT 1804 PHONE 308-2035 DATE 1/17/96

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and or relevant claim(s).

Search Seq. I. D. 3, 4 and 10 please

FOR OFFICIAL USE ONLY

14

STAFF USE ONLY

COMPLETED 1-30-96

SEARCHER

ONLINE TIME 90 TOTAL TIME 210
(in minutes)

NO. OF DATABASES 1

SYSTEMS
- ___ CAS ONLINE
- ___ DARC/QUESTEL
- ___ DIALOG
- ___ SDC
- ✓ OTHER IG

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:09:59 1996; MasPar time 3098.56 Seconds
1086.012 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-3
Description: (1:4448) from US08319745.seq
Perfect Score: 4448
N.A. Sequence: 1 GGCCAGTGTCGACACGCTGT.....CTTTTTTTTTTTTTTTTTTG 4448
Comp: CCGGTCACAGCTGTGCGACA.....GAAAAAAAAAAAAAAAAAAC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 478145 seqs, 378268764 bases x 2

Database: embl-new7
1 BCT
2 FUN
3 INV1
4 INV2
5 MAM
6 ORG
7 PLN
8 PRI
9 PRO
10 ROD
11 SYN
12 UNC
13 VRT
14 VIR

Database: EST-STS
15 EST1
16 EST2
17 EST3
18 EST4
19 EST5
20 EST6
21 EST7
22 EST8
23 EST9
24 EST10
25 EST11

26 EST12
27 EST13
28 EST14
29 EST15
30 EST16
31 EST17
32 EST18
33 EST19
34 EST20
35 EST21
36 EST22
37 EST23
38 EST24
39 EST25
40 EST26
41 EST27
42 EST28
43 EST29
44 EST30
45 EST31
46 EST32
47 EST33
48 EST34
49 EST35
50 EST36
51 EST37
52 EST38
53 EST39
54 EST40
55 EST41
56 EST42
57 EST43
58 EST44
59 EST45
60 EST46
61 EST47
62 EST48
63 EST49
64 EST50
65 EST51
66 EST52
67 EST53
68 EST54
69 EST55
70 EST56
71 EST57
72 EST58
73 STS1
74 STS2
75 STS3
76 STS4
77 gnEST1
78 gnEST2
79 gnEST3
80 gnEST4
81 gnEST5
82 gnEST6
83 gnEST7
84 gnEST8
85 gnEST9

Database: genbank89
86 BCT1
87 BCT2

88 BCT3
89 BCT4
90 BCT5
91 BCT6
92 INV1
93 INV2
94 INV3
95 INV4
96 INV5
97 MAM1
98 MAM2
99 PAT1
100 PAT2
101 PHG
102 PLN1
103 PLN2
104 PLN3
105 PLN4
106 PLN5
107 PLN6
108 PLN7
109 PRI1
110 PRI2
111 PRI3
112 PRI4
113 PRI5
114 PRI6
115 PRI7
116 PRI8
117 PRI9
118 ROD1
119 ROD2
120 ROD3
121 ROD4
122 ROD5
123 ROD6
124 ROD7
125 STR
126 SYN
127 UNA
128 VRL1
129 VRL2
130 VRL3
131 VRL4
132 VRL5
133 VRL6
134 VRT1
135 VRT2
136 VRT3

Database: genbank-new7
137 BCT
138 INV
139 MAM
140 PHG
141 PLN
142 PRI
143 ROD
144 STR
145 SYN
146 UNA
147 VRL
148 VRT

Database: u-embl43_89
149 part1

Statistics: Mean 12.432; Variance 3.452; scale 3.601

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	203	4.6	5536	93	DMPTCR	D. melanogaster patch	6.73e-241
	2	141	3.2	5665	94	DROMPP2	D.melanogaster membra	7.58e-151
	3	55	1.2	1301	116	HUMRDP	Human RD protein (RD)	3.59e-34
	4	55	1.2	1875	115	HUMMHRD4	Human MHC class III H	3.59e-34
	5	55	1.2	1226	111	HSRD	Human mRNA for RD pro	3.59e-34
	6	49	1.1	1805	103	DDIDPP5A	Dictyostelium purpure	5.73e-27
	7	47	1.1	1961	103	DDIDPP4A	Dictyostelium purpure	1.26e-24
	8	46	1.0	1831	136	XLU170KR	Xenopus laevis U1 70K	1.81e-23
	9	44	1.0	3963	93	DMU09306	Drosophila melanogast	3.54e-21
	10	44	1.0	255	57	T40389	ya33d03.r3 Homo sapie	3.54e-21
	11	43	1.0	6800	121	MUSIGMD	Mouse germline IgM ch	4.78e-20
	12	41	0.9	1332	136	XLU170K0	Xenopus laevis U1 70K	8.13e-18
	13	41	0.9	593	76	HUMUT1265	Human chromosome 2 ST	8.13e-18
	14	41	0.9	1812	121	MUSMHRDDK	Mouse MHC class III R	8.13e-18
	15	40	0.9	243	49	T02581	0224C3 Plasmodium fal	1.02e-16
	16	39	0.9	22703	93	CEZK675	Caenorhabditis elegan	1.25e-15
	17	39	0.9	22703	4	CEZK675	Caenorhabditis elegan	1.25e-15
	18	39	0.9	22703	138	CEZK675	Caenorhabditis elegan	1.25e-15
c	19	38	0.9	21805	4	CEF21H12	Caenorhabditis elegan	1.49e-14
c	20	38	0.9	21805	92	CELF21H12	Caenorhabditis elegan	1.49e-14
	21	38	0.9	486	65	T71726	yc62a01.r1 Homo sapie	1.49e-14
	22	37	0.8	1770	103	DDIDPP3A	Dictyostelium purpure	1.72e-13
	23	37	0.8	756	100	A25270	Human IFN-gamma antag	1.72e-13
	24	37	0.8	756	117	S74221	IK=IK factor [human,	1.72e-13
	25	37	0.8	756	8	HS221157	IK=IK factor [human,	1.72e-13
c	26	36	0.8	451	123	RATSAT	Rattus norvegicus (cl	1.92e-12
	27	36	0.8	215	22	HSC25B012	H. sapiens partial cD	1.92e-12
c	28	35	0.8	11026	122	MUSMVDPA	Mus musculus (clone l	2.08e-11
c	29	35	0.8	3408	123	RATSPR	Rat substance P recep	2.08e-11
	30	35	0.8	279	51	T15769	IB1866 Homo sapiens c	2.08e-11
	31	35	0.8	189	10	MMVIMV31	M.musculus DNA for vi	2.08e-11
	32	35	0.8	189	143	MMVIMV31	M.musculus DNA for vi	2.08e-11
	33	35	0.8	173	76	HUMUT266A	Human STS UT266, 5' p	2.08e-11
c	34	35	0.8	2552	123	RATSPR05	Rat substance P recep	2.08e-11
	35	34	0.8	471	98	U00207	Ovis aries subspecies	2.17e-10
	36	34	0.8	54670	122	RATCRYG	Rat gamma-crystallin	2.17e-10
	37	34	0.8	550	98	U00186	Capra aegagrus Saanen	2.17e-10
	38	34	0.8	434	34	R20917	yh18e08.r1 Homo sapie	2.17e-10
c	39	34	0.8	10272	120	MUSHOXMAA	Mouse Hox-3.1 gene an	2.17e-10
	40	34	0.8	479	98	U00206	Ovis aries Perendale	2.17e-10
c	41	33	0.7	3551	123	RNAALB	R.norvegicus (Sprague	2.19e-09
	42	33	0.7	8670	103	CRARG7	Chlamydomonas reinhar	2.19e-09
	43	33	0.7	3376	94	DROMPP1	D.melanogaster membra	2.19e-09
c	44	33	0.7	6125	119	MMU10098	Mus musculus 129/Sv c	2.19e-09
	45	33	0.7	6951	124	RNNMYC	R.norvegicus N-myc ge	2.19e-09

ALIGNMENTS

RESULT 1

```
LOCUS       DMPTCR       5536 bp    RNA             INV       09-APR-1991
DEFINITION  D. melanogaster patched (ptc) mRNA involved in segment polarity.
ACCESSION   X17558
KEYWORDS    developmental regulation; glycoprotein; patched gene;
            segmentation gene; transmembrane protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
            Metazoa; Eumetazoa; Bilateria; Coelomata; Protostomia; Arthropoda;
            Tracheata; Insecta; Pterygota; Neoptera; Holometabola; Diptera;
            Brachycera; Cyclorrhapha; Drosophilidae; Drosophila; Sophophora;
            melanogaster group; melanogaster subgroup.
REFERENCE   1  (bases 1 to 5536)
  AUTHORS   Ingham,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1990) to the EMBL/GenBank/DDBJ databases. Ingham
            P.W., Imperial Cancer Research Fund Developmental Biology Unit,
            Dept. of Zoology, South Parks Road, Oxford OX1 3PS, England
REFERENCE   2  (bases 1 to 5536)
  AUTHORS   Nakano,Y., Guerrero,I., Hidalgo,A., Taylor,A., Whittle,J.R. and
            Ingham,P.W.
  TITLE     A protein with several possible membrane-spanning domains encoded
            by the Drosophila segment polarity gene patched
  JOURNAL   Nature 341 (6242), 508-513 (1989)
  MEDLINE   90015164
COMMENT     See also M28418.
            See also <M28418> for related sequence.

            NCBI gi: 8389
FEATURES             Location/Qualifiers
     source          1..5536
                     /organism="Drosophila melanogaster"
                     /dev_stage="embryo"
                     /chromosome="2"
                     /map="44D"
     mRNA            1..5536
                     /gene="ptc"
     CDS             752..4651
                     /gene="ptc"
                     /standard_name="patched"
                     /note="NCBI gi: 8390"
                     /codon_start=1
                     /product="patched protein"
                     /translation="MDRDSLPRVPDTHGDVVDEKLFSDLYIRTSWVDAQVALDQIDKG
                     KARGSRTAIYLRSVFQSHLETLGSSVQKHAGKVLFVAILVLSTFCVGLKSAQIHSKVH
                     QLWIQEGGGLEAELAYTQKTIGEDESATHQLLIQTTHDPNASVLHPQALLAHLEVLVK
                     ATAVKVHLYDTEWGLRDMCNMPSTPSFEGIYYIEQILRHLIPCSIITPLDCFWEGSQL
                     LGPESAVVIPGLNQRLLWTTLNPASVMQYMKQKMSEEKISFDFETVEQYMKRAAIASG
                     YMEKPCLNPLNPNCPDTAPNKNSTQPPDVGAILSGGCYGYAAKHMHWPEELIVGGAKR
                     NRSGHLRKAQALQSVVQLMTEKEMYDQWQDNYKVHHLGWTQEKAAEVLNAWQRNFSRE
                     VEQLLRKQSRIATNYDIYVFSSAALDDILAKFSHPSALSIVIGVAVTVLYAFCTLLRW
                     RDPVRGQSSVGVAGVLLMCFSTAAGLGLSALLGIVFNAASTQVVPFLALGLGVDHIFM
                     LTAAYAESNRREQTKLILKKVGPSILFSACSTAGSFFAAAFIPVPALKVFCLQAAIVM
                     CSNLAAALLVFPAMISLDLRRRTAGRADIFCCCFPVWKEQPKVAPPVLPLNNNNGRGA
                     RHPKSCNNNRVALPAQNPLLEQRADIPGSSHSLASFSLATFAFQHYTPFLMRSWVKFL
                     TVMGFLAALISSLYASTRLQDGLDIIDLVPKDSNEHKFLDAQTRLFGFYSMYAVTQGN
                     FEYPTQQQLLRDYHDSFVRVPHVIKNDNGGLPDFWLLLFSEWLGNLQKIFDEEYRDGR
                     LTKECWFPNASSDAILAYKLIVQTGHVDNPVDKELVLTNRLVNSDGIINQRAFYNYLS
                     AWATNASSPTELLRANCIRNRANGASQGKLYPEPRQYFHQPNEYDLKIPKSLPLVYAQ
                     MPFYLHGLTDTSQIKTLIGHIRDLSVKYEGFGLPNYPSGIPFIFWEQYMTLRSSLAMI
                     LACVLLAALVLVSLLLLSVWAAVLVILSVLASLAQIFGAMTLLGIKLSAIPAVILILS
                     VGMMLCFNVLISLGFMTSVGNRQRRVQLSMQMSLGPLVHGMLTSGVAVFMLSTSPFEF
                     VIRHFCWLLLVVLCVGACNSLLVFPILLSMVGPEAELVPLEHPDRISTPSPLPVRSSK
                     RSGKSYVVQGSRSSRGSCQKSHHHHHKDLNDPSLTTITEEPQSWKSSNSSIQMPNDWT
                     YQPREQRPASYAAPPPAYHKAAAQQHHQHQGPPTTPPPPPFPTAYPPELQSIVVQPEVT
                     VETTHSDSNTTKVTATANIKVELAMPGRAVRSYNFTS"
     mat_peptide     752..4648
                     /gene="ptc"
                     /standard_name="patched"
                     /note="signal peptide sequence was not determined"
                     /product="patched protein"
     misc_feature    4896..4900
                     /gene="ptc"
                     /note="mRNA destabilizing motif"
     polyA_signal    4900..4905
                     /gene="ptc"
                     /note="atypical"
     misc_feature    5004..5008
                     /gene="ptc"
                     /note="mRNA destabilizing motif"
     misc_feature    5140..5144
                     /gene="ptc"
                     /note="mRNA destabilizing motif"
     misc_feature    5246..5250
                     /gene="ptc"
                     /note="mRNA destabilizing motif"
     misc_feature    5456..5460
                     /gene="ptc"
                     /note="mRNA destabilizing motif"
     polyA_signal    5503..5508
                     /gene="ptc"
     polyA_signal    5503..5508
                     /gene="ptc"
     polyA_site      5536
                     /gene="ptc"
BASE COUNT     1360 a   1575 c   1407 g   1194 t
ORIGIN
```

DB 93; Score 203; Match 59.9%; QryMatch 4.6%; Pred. No. 6.73e-241;
Matches 885; Conservative 0; Mismatches 574; Indels 18; Gaps 14;

```
Db    821 gatctttacatacgcaccagctgggtggacgcccaagtggcgctcgatcagatagataag 880
Qy    300 GATCTTTACATACGTACCAGTTGGGTGGACGCCCGCACTAGCTCTCTCTGAACTCGAAAAG 359

Db    881 ggcaaagcgcgtggcagccgcacggcgatctatctgcgatcagtattccagtcccacctc 940
Qy    360 GGTAACATCGAAGGAGGAAGAACCTCTCTGTGGATACGAGCGTGGCTACAAGAACAGCTC 419

Db    941 gaaaccctcggcagctccgtgcaaaagcacgcgggcaaggtgctattcgtggctatcctg 1000
Qy    420 TTTATTTTGGGCTGCTTTCTTCAAGGCGACGCGGGGAAAGTCCTCTTCGTTGCCATCCTC 479

Db   1001 gtgctgagcaccttctgcgtcggcctgaagagcgcccagatccactccaaggtgcaccag 1060
Qy    480 GTTCTGTCGACGTTCTGCGTCGGTCTCAAGTCAGCACAAATACATACAAGGGTCGACCAA 539

Db   1061 ctgtggatccaggagggcggcgggctggaggcggaactggcctacacacagaagacgatc 1120
Qy    540 CTCTGGGTTCAAGAGGGTGGTAGATTAGAAGCCGAGTTGAAATATACAGCGCAAGCTTTG 599

Db   1121 ggcgaggacgagtcggccacgcatcagctgctcattcagacgacccacgacccgaacgcc 1180
Qy    600 GGCGAGGCGGACTCCTCGACGCACCAGCTTGTCATACAAACTGCCAAAGATCCAGACGTC 659

Db   1181 tccgtcctgcatccgcaggcgctgcttgcccacctggaggtcctggtcaaggccaccgcc 1240
Qy    660 TCCCTGCTACATCCAGGCGCGTTGCTTGAACACCTTAAGGTGGTGCACGCAGCGACTCGG 719
```

```
Db   1241 gtcaaggtgcacctctacgacaccgaatgggggctgcgcgacatgtgcaacatgccgagc 1300
          || |  || ||| | |||||||  || ||| | ||    ||| ||||| |||  || |||
Qy    720 GTGACAGTTCACATGTACGACATTGAGTGGCGCCTCAAAGACCTGTGCTACAGCCCCAGC 779

Db   1301 acgccctccttcgagggcatctactacatcgagcagatcctgcgccacctcattccgtgc 1360
          |  ||   |||||||||   | || ||||||||   ||| |   | || |||| || |||
Qy    780 ATACCGGACTTCGAGGGTTACCACCACATCGAGTCAATCATAGACAACGTCATCCCCTGC 839

Db   1361 tcgatcatcacgccgctggactgtttctgggagggaagccagctgttgggtccggaatca 1420
          | || ||||| || || || || |||||||| ||   | || || | ||||| || |
Qy    840 GCTATTATCACCCCCCTTGATTGCTTCTGGGAAGGCTCCAAGTTGCTTGGTCCCGATTAT 899

Db   1421 gcggtcgttataccaggcctcaaccaacgactcctgtggaccacactgaatcccgcctct 1480
          |  |    |||||  || || ||  ||| | ||||| ||| | ||  ||   | |
Qy    900 CCTATATACGTACCACATCTTAAACACAAACTACAATGGAC-ACATTTAAATCCA--T-T 955

Db   1481 gtgatgcagtatatgaagcagaagatgtccgaggaaaagatcagcttcgacttcgagacc 1540
          | || |  ||| | ||||  ||| |  |   ||   | ||   |||  ||  |  ||
Qy    956 G-GAAGTTGTAGAAGAAGT-GAAAAAATTAAAGTTCCA-AT----TTCCTCTGAGC-ACA 1007

Db   1541 gtggagcagtacatgaagcgtgcggccattgcgagtggctacatggagaagccctgcctg 1600
          | |||  |||||||||| | || | |||  |   | ||||||| | ||||| ||| |
Qy   1008 ATAGAGGCGTACATGAAGAGAGCCGGCATCACTTCCGCCTACATGAAAAAGCCGTGCTTA 1067

Db   1601 aacccactgaatcccaattgcccggacacggcaccgaacaagaacagcacccagccgccg 1660
          ||||     | ||  |||| || | |||||| || ||||| ||        ||    ||
Qy   1068 GACCCCACCGACCCACATTGTCCAGCCACGGCTCCAAACAAAAAGTCTGGTCATATTCCA 1127

Db   1661 gatgtgggagccatcctgtccggaggctgctacggttatgccgcgaagcacatgcactgg 1720
          ||||| |  ||    |||||    || || || ||||  || ||     ||||||||||
Qy   1128 GATGTAGCGGCGGAGCTGTCGCACGGATGTTATGGTTTCGCGGCAGCTTACATGCACTGG 1187

Db   1721 ccggaggagctgattgtgggcggagcgaagaggaaccgcagcggacacttgaggaaggcc 1780
          |||||  || | ||||| || |||||| |  |||||    |    |    |||| || ||
Qy   1188 CCGGAACAGTTAATTGTAGGGGGAGCTACAAGGAATTCGACATCAGCTCTGAGAAAAGCA 1247

Db   1781 caggccctgcagtcggtggtgcagctgatgaccgagaaggaaatgtacgaccagtggcag 1840
          |  |   | ||| | || || ||| | |||  |||||  ||||||||||  | |||
Qy   1248 CGCGTT-TACAGACTGTAGTACAGTTAATGGGCGAGAGAGAAATGTACGAGTACTGGGCC 1306

Db   1841 gacaactacaaggtgcaccatcttggatggacgcaggagaaggcagcggaggttttgaac 1900
          ||  | || || || || ||  |||| ||||  || ||||||||||| |  ||  || |
Qy   1307 GATCATTATAAAGTACATCAAATTGGCTGGAATCAAGAGAAGGCAGCTGCTGTACTGGAT 1366

Db   1901 gcctggcagcgcaacttttc-gcgggaggtggaacagctgctacgtaaacagtcgagaat 1959
          ||||||||| | || ||| | || | ||   ||| |  | |||| | |  | || || ||
Qy   1367 GCCTGGCAGAGAAAGTTTGCCGCTGAAGTCAGAAAAATTACTACCTCAGGA-TC-AGTAT 1424

Db   1960 tgccaccaactacgatatctacgtgttcagctcggctgcactggatgacatcctggccaa 2019
             |  |   || | | ||||   ||||  | |  |  ||||  |||||||| || |  ||
Qy   1425 CATCGGCTTATA-GTT-TCTATCCGTTCTCCACCTCGACACTTAATGACATACTCGGGAA 1482

Db   2020 gttctcccatcccagcgccttgtccattgtcatcggcgtggccgtcaccgttttgtatgc 2079
          |||||||  |           |  |||| |  | |||       | |   |  | ||||
Qy   1483 GTTCTCCGAAGTGTCACTGAAGAACATTATATTAGGCTATATGTTTATGTTAATTTATGT 1542

Db   2080 cttctgcacgctcctccgctggagggaccccgtccgtggacagagcagtgtcggcgtggc 2139
             |   ||  |  | |  ||| |||| ||  | ||    ||     |||| ||  | ||
Qy   1543 TGCCGTTACTTTAATACAATGGCGGGATCCGATTCGCTCGCAAGCGGGTGTGGGTATCGC 1602

Db   2140 cggagttctgctcatgtgcttttagtaccgccgccggattgggattgtcagccctgctcgg 2199
          |||||||||  ||  | |  | | |  ||||| || || ||||| |  ||  |  | ||
Qy   1603 CGGAGTTCTACTACTATCAATCACTGTTGCCGCTGGCTTAGGATTTTGTGCTTTATTAGG 1662
```



```
Db   2200 tatcgttttcaatgccgccagcacccaggtggttccgtttttggcccttggtctgggcgt 2259
             ||    ||||| ||  | || || ||  | || || ||  | || || ||  | || ||
Qy   1663 CATACCATTCAACGCTTCAAGTACGCAAATAGTACCATTCCTAGCGCTCGGGTTAGGAGT 1722

Db   2260 cgatcacatcttcatgctgaccgctgcctatgcggag 2296
             |  | || ||  | || ||     | |||| ||||
Qy   1723 TCAAGATATGTTTCTTCTCACTCACACGTATGTGGAG 1759
```

```
RESULT    2
LOCUS       DROMPP2      5665 bp    DNA             INV       12-JAN-1993
DEFINITION  D.melanogaster membrane protein (patched) gene, complete cds.
ACCESSION   M28999
KEYWORDS    transmembrane protein.
SEGMENT     2 of 2
SOURCE      D.melanogaster (embryo), DNA and cDNA to mRNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Animalia; Metazoa; Arthropoda; Uniramia; Insecta;
            Pterygota; Neoptera; Holometabola; Diptera; Brachycera;
            Cyclorrhapha; Schizophora; Drosophiloidea; Drosophilidae.
REFERENCE   1  (bases 1 to 5665)
  AUTHORS   Hooper,J.E. and Scott,M.P.
  TITLE     The Drosophila patched gene encodes a putative membrane protein
            required for segmental patterning
  JOURNAL   Cell 59, 751-765 (1989)
  MEDLINE   90058658
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by J.E.Hooper, 27-SEP-1989.
            Mak called J.E.Hooper and requested copy of cds be sent showing
            introns and exons, 9-OCT-1989.  Copy was received and corrections
            made, 18-OCT-1989.

            NCBI gi: 552097
FEATURES             Location/Qualifiers
     source          1..5665
                     /organism="Drosophila melanogaster"
                     /sequenced_mol="DNA"
     CDS             join(M28418:3161..3289,95..648,986..2912,3051..3258,
                     3322..3593,3789..4559)
                     /gene="tuf"
                     /note="Description: tufted;  Allele:;  membrane protein;
                     NCBI gi: 552099"
                     /codon_start=1
                     /translation="MDRDSLPRVPDTHGDVVDEKLFSDLYIRTSWVDAQVALDQIDKG
                     KARGSRTAIYLRSVFQSHLETLGSSVQKHAGKVLFVAILVLSTFCVGLKSAQIHSKVH
                     QLWIQEGGRLEAELAYTQKTIGEDESATHQLLIQTTHDPNASVLHPQALLAHLEVLVK
                     ATAVKVHLYDTEWGLRDMCNMPSTPSFEGIYYIEQILRHLIPCSIITPLDCFWEGSQL
                     LGPESAVVIPGLNQRLLWTTLNPASVMQYMKQKMSEEKISFDFETVEQYMKRAAIGSG
                     YMEKPCLNPLNPNCPDTAPNKNSTQPPDVGAILSGGCYGYAAKHMHWPEELIVGGRKR
                     NRSGHLRKAQALQSVVQLMTEKEMYDQWQDNYKVHHLGWTQEKAAEVLNAWQRNFSRE
                     VEQLLRKQSRIATNYDIYVFSSAALDDILAKFSHPSALSIVIGVAVTVLYAFCTLLRW
                     RDPVRGQSSVGVAGVLLMCFSTAAGLGLSALLGIVFNAASTQVVPFLALGLGVDHIFM
                     LTAAYAESNRREQTKLILKKVGPSILFSACSTAGSFFAAAFIPVPALKVFCLQAAIVM
                     CSNLAAALLVFPAMISLDLRRRTAGRADIFCCCFPVWKEQPKVAPPVLPLNNNNGRGA
                     RHPKSCNNNRVPLPAQNPLLEQRADIPGSSHSLASFSLATFAFQHYTPFLMRSWVKFL
                     TVMGFLAALISSLYASTRLQDGLDIIDLVPKDSNEHKFLDAQTRLFGFYSMYAVTQGN
                     FEYPTQQQLLRDYHDSFVRVPHVIKNDNGGLPDFWLLLFSEWLGNLQKIFDEEYRDGR
                     LTKECWFPNASSDAILAYKLIVQTGHVDNPVDKELVLTNRLVNSDGIINQRAFYNYLS
                     AWATNDVFAYGASQGKLYPEPRQYFHQPNEYDLKIPKSLPLVYAQMPFYLHGLTDTSQ
                     IKTLIGHIRDLSVKYEGFGLPNYPSGIPFIFWEQYMTLRSSLAMILACVLLAALVLVS
                     LLLLSVWAAVLVILSVLASLAQIFGAMTLLGIKLSAIPAVILILSVGMMLCFNVLISL
                     GFMTSVGNRQRRVQLSMQMSLGPLVHGMLTSGVAVFMLSTSPFEFVIRHFCWLLLVVL
                     CVGACNSLLVFPILLSMVGPEAELVPLEHPDRISTPSPLPVRSSKRSGKSYVVQGSRS
```

```
                     SRGSCQKSHHHHHKDLNDPSLTTITEEPQSWKSSNSSIQMPNDWTYQPREQRPASYAA
                     PPPAYHKAAAQQHHQHQGPPTTPPPPFPTAYPPELQSIVVQPEVTVETTHSDSNTTKV
                     TATANIKVELAMPGRAVRSYNFTS"
     intron          <1..94
                     /note="MPR intron A; 9.2 kb gap in intron"
     exon            95..648
                     /gene="tuf"
                     /note="Description: tufted;  Allele: "
                     /number=2
     intron          649..985
                     /note="MPR intron B"
     exon            986..2912
                     /gene="tuf"
                     /note="Description: tufted;  Allele: "
                     /number=3
     intron          2913..3050
                     /note="MPR intron C"
     exon            3051..3258
                     /gene="tuf"
                     /note="Description: tufted;  Allele: "
                     /number=4
     intron          3259..3321
                     /note="MPR intron D"
     exon            3322..3593
                     /gene="tuf"
                     /note="Description: tufted;  Allele: "
                     /number=5
     intron          3594..3788
                     /note="MPR intron E"
     exon            3789..>4559
                     /gene="tuf"
                     /note="membrane protein;  Description: tufted;  Allele: "
                     /number=6
BASE COUNT     1397 a   1537 c   1350 g   1381 t
ORIGIN      About 9.2kb after segment 1, Chromosome 2, map position 44D-E.

  DB 94;  Score     141;  Match 65.1%;  QryMatch 3.2%;  Pred. No. 7.58e-151;
  Matches   304;  Conservative     0;  Mismatches 163;  Indels   0;  Gaps   0;

Db    164 ggcagctccgtgcaaaagcacgcgggcaaggtgctattcgtggctatcctggtgctgagc 223
          ||| |||   | |||   ||||||| || || || ||||| || ||||| || |||
Qy    429 GGCTGCTTTCTTCAAGGCGACGCGGGCAAAGTCCTCTTCGTTGCCATCCTCGTTCTGTCG 488

Db    224 accttctgcgtcggcctgaagagcgcccagatccactccaaggtgcaccagctgtggatc 283
          || |||||||||| || |||   || || || ||  | | |||  |||| || ||| |
Qy    489 ACGTTCTGCGTCGGTCTCAAGTCAGCACAAATACATACAAGGGTCGACCAACTCTGGGTT 548

Db    284 caggagggcggccggctggaggcggaactggcctacacacagaagacgatcggcgaggac 343
          || ||||| ||  |  | || || ||  ||   || ||| | |  |  | |||||||
Qy    549 CAAGAGGGTGGTAGATTAGAAGCCGAGTTGAAATATACAGCGCAAGCTTTGGGCGAGGCG 608

Db    344 gagtcggccacgcatcagctgctcattcagacgacccacgacccgaacgcctccgtcctg 403
          || ||  | ||||| |||||  |||| || ||  || | || ||  ||| |||| | ||
Qy    609 GACTCCTCGACGCACCAGCTTGTCATACAAACTGCCAAAGATCCAGACGTCTCCCTGCTA 668

Db    404 catccgcaggcgctgcttgcccacctggaggtcctggtcaaggccaccgccgtcaaggtg 463
          |||||    ||| ||||||  |||||  ||||  ||  |   || ||    || |  ||
Qy    669 CATCCAGGCGCGTTGCTTGAACACCTTAAGGTGGTGCACGCAGCGACTCGGGTGACAGTT 728

Db    464 cacctctacgacaccgaatgggggctgcgcgacatgtgcaacatgccgagcacgccctcc 523
          ||| | |||||||  || ||| | ||    ||| ||||| ||| || ||||  ||   |
Qy    729 CACATGTACGACATTGAGTGGCGCCTCAAAGACCTGTGCTACAGCCCCAGCATACCGGAC 788

Db    524 ttcgagggcatctactacatcgagcagatcctgcgccacctcattccgtgctcgatcatc 583
```



```
          |||||||   | || |||||||   ||| |   | || |||| || ||| | || |||
Qy    789 TTCGAGGGTTACCACCACATCGAGTCAATCATAGACAACGTCATCCCCTGCGCTATTATC 848

Db    584 acgccgctggactgtttctgggagggaagccagctgttgggtccgga 630
          || || || || || |||||||| ||   | || || | ||||| ||
Qy    849 ACCCCCCTTGATTGCTTCTGGGAAGGCTCCAAGTTGCTTGGTCCCGA 895


RESULT    3
LOCUS       HUMRDP       1301 bp    mRNA            PRI       28-SEP-1993
DEFINITION  Human RD protein (RD) mRNA, complete cds.
ACCESSION   L03411
KEYWORDS    RD protein.
SOURCE      Homo sapiens (library: HL1016b; Clontech) cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 1301)
  AUTHORS   Cheng,J., Macon,K.J. and Volanakis,J.E.
  TITLE     cDNA cloning and characterization of the protein encoded by RD, a
            gene located in the class III region of the human major
            histocompatibility complex
  JOURNAL   Biochem. J. 294 ( Pt 2), 589-593 (1993)
  MEDLINE   93384537
COMMENT     NCBI gi: 190973
FEATURES             Location/Qualifiers
     source          1..1301
                     /organism="Homo sapiens"
                     /cell_line="Jurkat"
                     /cell_type="T-cell"
                     /sequenced_mol="cDNA to mRNA"
                     /tissue_lib="HL1016b; Clontech"
     CDS             87..1229
                     /gene="RD"
                     /map="1p36-p22.1"
                     /note="located in the MHC class III region;  NCBI gi:
                     190974"
                     /codon_start=1
                     /product="RD protein"
                     /translation="MLVIPPGLSEEEEALQKKFNKLKKKKKALLALKKQSSSSTTSQG
                     GVKRSLSEQPVMDTATATEQAKQLVKSGAISAIKAETKNSGFKRSRTLEGKLKDPEKG
                     PVPTFQPFQRSISADDDLQESSRRPQRKSLYESFVSSSDRLRELGPDGEEAEGPGAGD
                     GPPRSFDWGYEERSGAHSSASPPRSRSRDRSHERNRDRDRDRERDRDRDRDRDRERDR
                     DRDRDRDRDRERDRDRDRDREGPFRRSDSFPERRAPRKGNTLYVYGEDMTPTLL
                     RGAFSPFGNIIDLSMDPPRNCAFVTYEKMESADQAVAELNGTQVESVQLKVNIARKQP
                     MLDAATGKSVWGSLAVQNSPKGCHRDKRTQIVYSDDVYKENLVDGF"
BASE COUNT      337 a    353 c    388 g    223 t
ORIGIN

  DB 116;  Score      55;  Match 73.5%;  QryMatch 1.2%;  Pred. No. 3.59e-34;
  Matches    86;  Conservative     0;  Mismatches  31;  Indels   0;  Gaps   0;

Db    687 gatcgggagcgggatcgagaccgggatcgagacagagacagagagcgggacagggatcgg 746
          ||||| ||    ||| | || || || || || || |||||||| || ||||||||||||
Qy   3951 GATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGG 4010

Db    747 gatcgggatcgagatcgagaccgggaacgggacagggatcgggagcgggatcgagac 803
          ||| ||||||| || || ||| ||||  |||| || || ||   | | || ||||||
Qy   4011 GATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGAC 4067


RESULT    4
LOCUS       HUMMHRD4     1875 bp    DNA             PRI       15-JUN-1990
DEFINITION  Human MHC class III HLA-RD gene, exons 5 through 10.
```

```
ACCESSION   M33230 M30115
KEYWORDS    class III gene; major histocompatibility complex.
SEGMENT     4 of 5
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 1875)
  AUTHORS   Speiser,P.W. and White,P.C.
  TITLE     Structure of the human RD gene: A highly conserved gene in the
            class III region of the major histocompatibility complex
  JOURNAL   DNA 8, 745-751 (1989)
  MEDLINE   90126228
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by P.C.White, 21-NOV-1989.

            NCBI gi: 188510
FEATURES             Location/Qualifiers
     source          1..1875
                     /organism="Homo sapiens"
     intron          <1..12
                     /note="MHC HLA-RD intron D"
     exon            13..90
                     /number=5
     intron          91..212
                     /note="MHC HLA-RD intron E"
     exon            213..250
                     /number=6
     intron          251..417
                     /note="MHC HLA-RD intron F"
     exon            418..755
                     /number=7
     intron          756..865
                     /note="MHC HLA-RD intron G"
     exon            866..1013
                     /number=8
     intron          1014..1182
                     /note="MHC HLA-RD intron H"
     exon            1183..1237
                     /number=9
     intron          1238..1487
                     /note="MHC HLA-RD intron I"
     exon            1488..1590
                     /number=10
     intron          1591..>1875
                     /note="MHC HLA-RD intron J"
BASE COUNT      436 a    479 c    511 g    449 t
ORIGIN      About 1.4 kb after segment 3.

  DB 115;  Score      55;  Match 73.5%;  QryMatch 1.2%;  Pred. No. 3.59e-34;
  Matches    86;  Conservative    0;  Mismatches  31;  Indels   0;  Gaps   0;

Db    614 gatcgggagcgggatcgagaccgggatcgagacagagacagagagcgggacagggatcgg 673
          ||||| ||    ||| | || || || || || || ||||||||| || ||||||||||||
Qy   3951 GATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGG 4010

Db    674 gatcgggatcgagatcgagaccgggaacgggacagggatcgggagcgggatcgagac 730
          ||| |||||||| || || ||| ||||  |||| || || ||   | | || ||||||
Qy   4011 GATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGAC 4067


RESULT    5
LOCUS       HSRD         1226 bp    RNA            PRI       12-SEP-1993
DEFINITION  Human mRNA for RD protein, RNA-binding.
```



```
ACCESSION   X16105
KEYWORDS    RNA binding protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
            Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1226)
  AUTHORS   Spritz,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1989) to the EMBL/GenBank/DDBJ databases. Spritz
            R.A., University of Wisconsin-Madison, 309 Laboratory of Genetics,
            445 Henry Mall, Madison WI 53706, U S A
REFERENCE   2  (bases 1 to 1226)
  AUTHORS   Surowy,C.S., Hoganson,G., Gosink,J., Strunk,K. and Spritz,R.A.
  TITLE     The human RD protein is closely related to nuclear RNA-binding
            proteins and has been highly conserved
  JOURNAL   Gene 90 (2), 299-302 (1990)
  MEDLINE   90382680
COMMENT     A second cDNA sequence (clone hRD2 from lgt11 placenta library)
            corresponds to bases 1-1082 and is identical.

            NCBI gi: 35912
FEATURES             Location/Qualifiers
     source          1..1226
                     /organism="Homo sapiens"
                     /clone_lib="lgt11 liver"
                     /clone="hRD1"
                     /chromosome="6"
     CDS             36..1013
                     /note="RD protein (AA 1-325);  NCBI gi: 35913"
                     /codon_start=1
                     /translation="MDTATATEQAKQLVKSGAISAIKAETKNSGFKRSRTLEGKLKDP
                     EKGPVPTFQPFQRSISADDDLQESSRRPQRKSLYESFVSSSDRLRELGPDGEEAEGPG
                     AGDGPPRSFDWGYEERMGAHSSASPPRSRSRDRSHERNRDRDRDRERDRDRDRDRDRE
                     RDRDRDRDRDRDRERDRDRERDRDRDREGPFRRSDSFPERRAPRKGNTLYVYGEDMTP
                     TLLRGAFSPFGNIIDLSMDPPRNCAFVTYEKMESADQAVAELNGTQVESVQLKVNIAR
                     KQPMLDAATGKSVWGSLAVQNSPKGCHRDKRTQIVYSDDVYKENLVDGF"
     polyA_site      1226
                     /note="polyadenylation site"
BASE COUNT      311 a    333 c    348 g    234 t
ORIGIN

  DB 111;  Score      55;  Match 73.5%;  QryMatch 1.2%;  Pred. No. 3.59e-34;
  Matches    86;  Conservative    0;  Mismatches  31;  Indels   0;  Gaps   0;

Db    471 gatcgggagcgggatcgagaccgggatcgagacagagacagagagcgggacagggatcgg 530
          ||||| ||    ||| | || || || || || || ||||||||| || ||||||||||||
Qy   3951 GATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGG 4010

Db    531 gatcgggatcgagatcgagaccgggaacgggacagggatcgggagcgggatcgagac 587
          ||| |||||||| || || ||| ||||  |||| || || ||   | | || ||||||
Qy   4011 GATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGAC 4067


RESULT    6
LOCUS       DDIDPP5A     1805 bp    DNA            PLN       11-MAR-1994
DEFINITION  Dictyostelium purpureum (Dpp5) DNA sequence, repeat region.
ACCESSION   L05617
KEYWORDS    repeat region.
SOURCE      Dictyostelium purpureum (strain DPA) DNA.
  ORGANISM  Dictyostelium purpureum
```

```
            Eukaryota; Fungi; Myxomycota; Acrasiomycetes; Dictyosteliales;
            Dictyosteliaceae.
REFERENCE   1  (bases 1 to 1805)
  AUTHORS   Kiyosawa,H., Hughes,J.E., Podgorski,G.J. and Welker,D.L.
  TITLE     Small circular plasmids of the eukaryote Dictyostelium purpureum
            define two novel plasmid families
  JOURNAL   Plasmid 30, 106-118 (1993)
  MEDLINE   94052367
COMMENT     NCBI gi: 167774
FEATURES             Location/Qualifiers
     source          1..1805
                     /organism="Dictyostelium purpureum"
                     /strain="DPA"
                     /sequenced_mol="DNA"
     repeat_region   320..477
BASE COUNT      783 a    159 c    202 g    661 t
ORIGIN

  DB 103;  Score      49;  Match 72.5%;  QryMatch 1.1%;  Pred. No. 5.73e-27;
  Matches    87;  Conservative    0;  Mismatches  32;  Indels   1;  Gaps   1;

Db    319 gatcgagaaagagaaagagacagagatagagacagagatagagatagagatagagaaaga 378
          ||||||||  | |||||  ||||||||| | |||||  ||| | |||||  ||| | ||  |
Qy   3969 GATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGGGATCGTGACCGG 4028

Db    379 gaaagagaaagagatagagacagagatcgagaaagagaaagagagagatagagacagaga 438
          || || || || |||||||||  ||    |||||  ||||  || | ||||  || | |||||
Qy   4029 GACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG-GCGAGACCGATATAGAGA 4087


RESULT    7
LOCUS       DDIDPP4A     1961 bp    DNA             PLN       11-MAR-1994
DEFINITION  Dictyostelium purpureum (Dpp4) DNA sequence, repeat region.
ACCESSION   L05616
KEYWORDS    repeat region.
SOURCE      Dictyostelium purpureum (strain V15) DNA.
  ORGANISM  Dictyostelium purpureum
            Eukaryota; Fungi; Myxomycota; Acrasiomycetes; Dictyosteliales;
            Dictyosteliaceae.
REFERENCE   1  (bases 1 to 1961)
  AUTHORS   Kiyosawa,H., Hughes,J.E., Podgorski,G.J. and Welker,D.L.
  TITLE     Small circular plasmids of the eukaryote Dictyostelium purpureum
            define two novel plasmid families
  JOURNAL   Plasmid 30, 106-118 (1993)
  MEDLINE   94052367
COMMENT     NCBI gi: 167773
FEATURES             Location/Qualifiers
     source          1..1961
                     /organism="Dictyostelium purpureum"
                     /strain="V15"
                     /sequenced_mol="DNA"
     repeat_region   320..479
BASE COUNT      877 a    161 c    258 g    665 t
ORIGIN

  DB 103;  Score      47;  Match 73.3%;  QryMatch 1.1%;  Pred. No. 1.26e-24;
  Matches    74;  Conservative    0;  Mismatches  27;  Indels   0;  Gaps   0;

Db    319 gatcgagaaagagaaagagacagagatagagacagagatagagacagagatagagaaaga 378
          ||||||||  | |||||  ||||||||| | |||||  ||| | ||  || ||| | ||  |
Qy   3969 GATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGGGATCGTGACCGG 4028

Db    379 gaaagagaaagagaaagagaaagagaaagagaaagagaaagagaaag 419
          || || || || || ||||||  ||   ||||||  ||||  ||
```



```
Qy   4029 GACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG 4069


RESULT    8
LOCUS       XLU170KR     1831 bp    RNA             VRT       12-SEP-1993
DEFINITION  Xenopus laevis U1 70K mRNA.
ACCESSION   X12430
KEYWORDS    U1 small nuclear RNA.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
            Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amphibia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1831)
  AUTHORS   Etzerodt,M., Vignali,R., Scherly,D., Mattaj,I.W., Ciliberto,G. and
            Philipson,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1988) to the EMBL/GenBank/DDBJ databases.
            Etzerodt M., University of Aarhus, Dept. of Molecular Biology, C.F.
            Moellers Allee 130, DK-8000 Aarhus C, Denmark
REFERENCE   2  (bases 1 to 1831)
  AUTHORS   Etzerodt,M., Vignali,R., Ciliberto,G., Scherly,D., Mattaj,I.W. and
            Philipson,L.
  TITLE     Structure and expression of a Xenopus gene encoding an snRNP
            protein (U1 70K)
  JOURNAL   EMBO J. 7 (13), 4311-4321 (1988)
  MEDLINE   89210819
COMMENT     NCBI gi: 65178
FEATURES             Location/Qualifiers
     source          1..1831
                     /organism="Xenopus laevis"
                     /clone_lib="NM1149"
                     /clone="XcU170A"
     exon            <1..89
                     /number=1
     exon            90..246
                     /number=2
     CDS             99..1514
                     /note="NCBI gi: 65179"
                     /codon_start=1
                     /product="U1 70K"
                     /translation="MTQFLPPNLLALFAPRDPVPYLPPLDKLPHEKHHNQPYCGIAPY
                     IREFEDPRDAPPPTRAETREERMERKRREKIERRQQDVENELKIWDPHNDQNAQGDAF
                     KTLFVARVNYDTTESKLRREFEVYGPIKRIHIVYNKGSEGSGKPRGYAFIEYEHERDM
                     HSAYKHADGKKIDGRRVLVDVERGRTVKGWRPRRLGGGLGGTRRGGADVNIRHSGRDD
                     TSRYDERDRERERDRRERSRERERKEPRERRRSRSRERRRKSRSREKEERKRTREKSKD
                     KDKEKDKDNKDRDRKRRSRSRERKRERDRDREKKEERVEAEVPEADDAPQDDAQIGDL
                     GIDGIELKQEPEEKSRERDRERDRDREKGEKDRDKDRDRDRDRRRSHRDRDREKDRDR
                     DRDRRRDRDRDRERDKDHKRERDRGDRSEKREERVPDNGMVMEQAEETSQDMYLDQES
                     MQSGDGYLSTENGYMMEPPME"
     exon            247..308
                     /number=3
     exon            309..363
                     /number=4
     exon            364..428
                     /number=5
     exon            429..491
                     /number=6
     exon            492..582
                     /number=7
     misc_feature    537..560
                     /note="RNP consensus signal"
```

```
     exon            583..684
                     /number=8
     exon            685..772
                     /number=9
     exon            773..1831
                     /number=10
     polyA_signal    1810..1817
     polyA_site      1831
BASE COUNT      602 a    335 c    524 g    370 t
ORIGIN

  DB 136;  Score      46;  Match 65.3%;  QryMatch 1.0%;  Pred. No. 1.81e-23;
  Matches    98;  Conservative    0;  Mismatches  52;  Indels   0;  Gaps   0;

Db   1228 gaaggcgcagccatagagacagggatagagagaaggatagggacagagacagggacagga 1287
           |||||    || ||  ||| ||||||  ||||    || ||||||||||  | ||||||
Qy   3946 GAAGGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGG 4005

Db   1288 ggcgagatagggaccgcgatagagagagggacaaggatcacaaacgagagcgagacagag 1347
            || ||||||||  || ||  | || ||||| | ||||    |||||  | ||||  |||
Qy   4006 ATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAG 4065

Db   1348 gtgaccgcagcgaaaagagagaggagaggg 1377
               ||   |  | | ||||| || ||||
Qy   4066 ACAGGCGAGACCGATATAGAGACGAAAGGG 4095


RESULT    9
LOCUS       DMU09306     3963 bp    mRNA            INV       15-MAY-1994
DEFINITION  Drosophila melanogaster shuttle craft protein (stc) mRNA, complete
            cds.
ACCESSION   U09306
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eucaryotae; Metazoa; Arthropoda; Tracheata; Insecta; Pterygota;
            Diptera; Brachycera; Cyclorrhapha; Drosophilidae; Drosophila;
            (Sophophora); melanogaster group; melanogaster subgroup.
REFERENCE   1  (bases 1 to 3963)
  AUTHORS   Stroumbakis,N.D., Li,Z. and Tolias,P.P.
  TITLE     Dynamic subcellular distribution in egg chambers and embryos of a
            novel Drosophila melanogaster protein encoded by the embryonic
            lethal gene shuttle craft
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3963)
  AUTHORS   Tolias,P.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1994) Peter P. Tolias, Public Health Research
            Institute, 455 First Ave., New York, NY 10016, USA
COMMENT     NCBI gi: 487399
FEATURES             Location/Qualifiers
     source          1..3963
                     /clone_lib="ovarian gt22A of P. Tolias, ovarian gt11 of
                     L.Kalfayan and st.10 of A. Spradling"
                     /chromosome="2L"
                     /organism="Drosophila melanogaster"
                     /map="35C1-C3"
                     /sex="female"
                     /tissue_type="ovary"
                     /dev_stage="adult"
                     /note="Several overlapping clones used to assemble
                     sequence of shuttle craft/l(2)35Cb. Originally isolated as
                     a partial clone (containing the C-terminal 222 amino
                     acids) in expression cloning screens for cDNAs encoding
```



```
                     single stranded nucleic acid binding proteins"
     5'UTR           1..223
                     /map="35C1-C3"
     CDS             224..3544
                     /gene="stc"
                     /standard_name="shuttle craft"
                     /note="C-terminal 222 amino acids encode a novel
                     single-stranded DNA binding domain;  NCBI gi: 487400"
                     /codon_start=1
                     /evidence=experimental
                     /product="shuttle craft protein"
                     /translation="MAEYWQQLTNGPGEAGPGNESSAMVDCNGGHESAAVGGSCNRHS
                     NNNYVNFNQFIMQHNLGGGAPSNATSTMQHPVGSSYTNFSLGGGGGAFGLNPPVASAS
                     TSHFANVSHQSPNFYSQAMIPTYQNGDGIARVTVTSSYGSVNPSNSNFSSFYTPFGNN
                     PFDFSASKLQASAPEFVPNFAKLSLEETPAAATTNGNSTASLETAINETRPRTLRAQE
                     PAERGANNQCSNHNYERERERERDRDRDRERDRDRDRDRDRDRDRDRDRDRDRDRSRPGNTRQ
                     QRRSDYRDDREDRYERSDRRRPQKQQRYDNHRSNKRRDDWNRNRDRINGFPRAVDDLD
                     TSNESAHPSPEKQSQLQQISPRRGPPLPPADNEKLSQREKLVRDIEQRRLECLVCVEA
                     IKSHQPTWSCRNCYHMLHLKCTITWASSSKSEVGWRCPACQNVLQDLPRDYLCFCGKL
                     KNPPVSRTELAHSCGEVCCRIEGCSHACTLLCHPGPCPPCQANVVRSCGCGRSTKTMQ
                     CAMKEEVLCGEICDKLLNCGEHRCQAECHSGKCAACSEQVVQQCHCGKQERKVPCTRE
                     SQDKRTYSCKDSCGQPLPCGHHKCKDSCHAGSCRPCKLSPEQITSCPCGKIPVPAGQR
                     SSCLDPIPTCEGICSRTLRCGKPAHPHQCGSKCHLGQCPPCPKQTGVKCRCGHMDQMI
                     KCRQLCNRADDARCKRRCTKKRSCGKHKCNVECCIDIDHDCPLPCNRTLSCGKHKCDQ
                     PCHRGNCPPCYRSSFEELYCECGAEVIYPPVPCGTKKPICKLPSSRIHPCDHPPQHNC
                     HSGPTCPPCMIFTTKLCHGNHDWRKTIPCSQPNFSCGMACGKPLPCGGHKCIKPCHEG
                     PCQSAGEICRQSCTKPRPTCGHKCAAACHEGACPETPCKELVEVQCECGNRKQNRSCQ
                     ELAREHSRIATIQLASSMAEMSRGNYMELSEILAPAKKSNKTLDCNDECRLLERNRRL
                     AAALSSGNSDTKQKCLTKYSEFVRGFAKKNPALTKSVYETLTDLVKLAKESKQRSRSH
                     SFPTMNREKRQLVHELCEVFGIESVSYDKEPNRNVVATAHKDRCWFPATSIMEVLARE
                     SGQRRVPVPSNNAWGLKK"
     variation       replace(548..568,"")
                     /gene="stc"
                     /note="21 bp from 548..568 are not present in some cDNAs
                     due to alternative splicing. This results in an in-frame 7
                     amino acid deletion of the stc protein."
     repeat_region   923..1027
                     /gene="stc"
                     /note="RD-domain; a repeat region of R followed by D or E"
     misc_feature    1109..1159
                     /gene="stc"
                     /note="bipartite nuclear localization site (putative)"
     repeat_region   1967..2005
                     /gene="stc"
                     /note="PCC repeat (PXXXXXXXCXXXC)"
     repeat_region   2699..2737
                     /gene="stc"
                     /note="PCC repeat (PXXXXXXXCXXXC)"
     repeat_region   2744..2782
                     /gene="stc"
                     /note="PCC repeat (PXXXXXXXCXXXC)"
     repeat_region   2792..2830
                     /gene="stc"
                     /note="PCC repeat (PXXXXXXXCXXXC)"
     repeat_region   2837..2875
                     /gene="stc"
                     /note="PCC repeat (PXXXXXXXCXXXC)"
     3'UTR           3545..3963
                     /map="35C1-C3"
     polyA_signal    3937..3942
                     /map="35C1-C3"
     polyA_site      3963
                     /map="35C1-C3"
BASE COUNT     1043 a   1055 c   1111 g    754 t
```

```
ORIGIN

  DB 93;  Score      44;  Match 72.9%;  QryMatch 1.0%;  Pred. No. 3.54e-21;
  Matches    70;  Conservative    0;  Mismatches  26;  Indels   0;  Gaps   0;

Db    923 gaacgggagcgtgagcgggaacgtgaccgggaccgggacagggaaagagaccgggacaga 982
          || || ||  | ||  | || || ||| |||| ||||| |||||  | ||||||||||||
Qy   3975 GACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGGGATCGTGACCGGGACAGG 4034

Db    983 gacagggacagagaccgggacagagaccgggacagg 1018
          || ||||| ||||| ||    ||||| || ||||||
Qy   4035 GATAGGGATAGAGAACGATCGAGAGAACGAGACAGG 4070


RESULT   10
LOCUS       T40389        255 bp    mRNA            EST       08-FEB-1995
DEFINITION  ya33d03.r3 Homo sapiens cDNA clone 62405 5' contains TAR1
            repetitive element,.
ACCESSION   T40389
KEYWORDS    EST.
SOURCE      human clone=62405 library=Stratagene ovary (#937217)
            vector=Bluescript SK host=SOLR cells (kanamycin resistant)
            primer=M13RP1 Rsite1=EcoRI Rsite2=XhoI Cloned unidirectionally.
            Primer: Oligo dT. Total ovary tissue, normal 49 year old caucasian
            female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor
            sequence: 5'-GAATTCGGCACGAG-3'; 3' adaptor sequence:
            5'-CTCGAGTTTTTTTTTTTTTTTTT-3'.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 255)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

            NCBI gi: 648017
FEATURES             Location/Qualifiers
     source          1..255
                     /organism="Homo sapiens"
                     /clone="62405"
                     /note="human"
BASE COUNT      112 a     30 c     97 g     16 t
ORIGIN

  DB 57;  Score      44;  Match 73.9%;  QryMatch 1.0%;  Pred. No. 3.54e-21;
  Matches    68;  Conservative    0;  Mismatches  24;  Indels   0;  Gaps   0;

Db    141 gggaacgagaacgggatagggaccgtgaccggacaaaagagagagaccgagatcgggatc 200
          |||| || ||    |||||||| || |||||   ||  || ||||| || ||  ||||||
Qy   3949 GGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATC 4008
```



```
Db    201 gagagagagatcgtgaccgggatagagaaagg 232
          | || || ||||||||||||||| || || |||
Qy   4009 GGGATAGGGATCGTGACCGGGACAGGGATAGG 4040


RESULT   11
LOCUS       MUSIGMD      6800 bp    DNA             ROD       28-FEB-1994
DEFINITION  Mouse germline IgM chain gene, mu-delta region.
ACCESSION   K02138 M65580 M65581 M65582 M65583 M65584 M65585 X00142
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin mu-chain; immunoglobulin-delta;
            unidentified reading frame.
SOURCE      Mouse DNA, clones PJS5, PCP18, PGG-mu-9.
  ORGANISM  Mus musculus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (sites)
  AUTHORS   Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and
            Hood,L.E.
  TITLE     Two mRNAs can be produced from a single immunoglobulin mu gene by
            alternative RNA processing pathways
  JOURNAL   Cell 20, 313-319 (1980)
  MEDLINE   80222874
REFERENCE   2  (sites)
  AUTHORS   Tucker,P.W., Liu,C.P., Mushinski,J.F. and Blattner,F.R.
  TITLE     Mouse immunoglobulin D: Messenger RNA and genomic DNA sequences
  JOURNAL   Science 209, 1353-1360 (1980)
  MEDLINE   81015444
REFERENCE   3  (bases 1 to 6800)
  AUTHORS   Richards,J.E., Gilliam,A.C., Shen,A., Tucker,P.W. and Blattner,F.R.
  TITLE     Unusual sequences in the murine immunoglobulin m-delta heavy-chain
            region
  JOURNAL   Nature 306, 483-487 (1983)
  MEDLINE   84068187
COMMENT     Cell 20, 313-319 (1980) represents bases 1709-2428
             Science 209, 1353-1360 (1980) represents bases  4631 to 4934; 5311
            to 5415; 6415 to 6735; 6736 to 6800
             Nature 306, 483-487 (1983) represents bases 1 to 6800. There is an
            open reading (ORF-146) at bases 3460-3901.  There are several
            stretches which may form Z-DNA [3].  The deletion in myeloma TEPC
            1017 (see separate entry) ends at base 4395.  [3] also notes the
            presence of a possible C-delta-2 pseudogene just upstream of the
            C-delta-3 exon.

            NCBI gi: 202416
FEATURES             Location/Qualifiers
     source          1..6800
                     /organism="Mus musculus"
                     /germline
                     /sequenced_mol="DNA"
     intron          <1
                     /note="C-mu-secreted intron"
     intron          <1..4652
                     /note="C-delta J-C intron"
     intron          <1..1862
                     /note="C-mu-membrane intron"
     CDS             <2..63
                     /note="C-mu-secreted, last;  NCBI gi: 457143"
                     /codon_start=2
                     /product="immunoglobulin mu-chain"
                     /translation="KPTLYNVSLIMSDTGGTCY"
     misc_feature    108..743
                     /note="z-DNA region; putative"
```

```
     polyA_signal    163..168
                     /note="for Cmu(s)"
     polyA_signal    676..681
                     /note="putative"
     repeat_region   1097..2096
                     /note="(CA)33 repeat"
                     /rpt_unit=1097..1098
     exon            1863..1978
                     /note="IgH chain C-mu-membrane, exon m1"
     CDS             join(1863..1978,2097..2105)
                     /partial
                     /note="C-mu-membrane;  NCBI gi: 202417"
                     /codon_start=3
                     /product="immunoglobulin mu-chain"
                     /translation="GEVNAEEEGFENLWTTASTFIVLFLLSLFYSTTVTLFKVK"
     intron          1979..2096
                     /note="C-mu-membrane intron A"
     exon            2097..2105
                     /note="Ig H-chain C-mu-membrane, exon m2"
     polyA_signal    2348..2353
                     /note="for Cmu (membrane); putative"
     repeat_region   2556..2627
                     /note="(GGGAGA)12-repeat"
                     /rpt_unit=2556..2561
     repeat_region   2636..2691
                     /note="(GA)28-repeat"
                     /rpt_unit=2636..2638
     polyA_signal    3367..3372
                     /note="for Cmu (membrane); putative"
     promoter        3367..3372
                     /note="TATA-like motive for ORF 146; putative"
     CDS             3479..3901
                     /note="ORF 146 with stretch of 27 hydrophobic amino acids,
                     membrane binding segment; putative;  NCBI gi: 457144"
                     /codon_start=1
                     /translation="MIGLVTKFQIYCSHSPWYGTSHTSIYFCYPGMLGQSATHGSLSS
                     RTLFFPSSRGQKCGPQQALLCPQPTILGWCCFKLVIWVLSQSSEAIAVGQSLSPGTNI
                     PVTVCPLYLMPCFKNPVLLGLMPTLIQHKFISTRFHPP"
     repeat_unit     3509..3680
                     /note="unique sequence inverted repeat 1"
     repeat_unit     3936..4098
                     /note="unique sequence inverted repeat 1"
     CDS             join(4653..4934,5311..5415,6415..6735)
                     /partial
                     /note="C-delta, CH1 domain;  NCBI gi: 554163"
                     /codon_start=3
                     /product="immunoglobulin mu-chain"
                     /translation="DKKEPDMFLLSECKAPEENEKINLGCLVIGSQPLKISWEPKKSS
                     IVEHVFPSEMRNGNYTMVLQVTVLASELNLNHTCTINKPKRKEKPFKFPESWDSQSSK
                     RVTPTLQAKNHSTEATKAITTKKDIEGAMAPSNLTVNILTTSTHPEMSSWLLCEVSGF
                     FPENIHLMWLGVHSKMKSTNFVTANPTAQPGGTFQTWSVLRLPVALSSSLDTYTCVVE
                     HEASKTKLNASKSLAIS"
     exon            4653..4934
     CDS             4653..4934
                     /pseudo
                     /codon_start=1
     polyA_signal    4895..4900
                     /note="for ORF 146; putative"
     intron          4935..5310
                     /note="C-delta intron A"
     exon            5311..5415
                     /partial
                     /note="Ig H-chain C-delta, hinge domain"
     intron          5416..6414
```



```
                     /note="C-delta intron B"
     repeat_region   5851..5910
                     /note="(CT)30-repeat"
                     /rpt_unit=5851..5852
     repeat_region   5911..5970
                     /note="(CA)30-repeat"
     CDS             5989..6325
                     /note="C gamma 3; putative"
                     /pseudo
                     /codon_start=1
     exon            6415..6735
                     /note="Ig H-chain C-delta, CH3 domain"
     intron          6736..>6800
                     /note="C-delta intron C"
BASE COUNT     1953 a    1758 c    1500 g    1589 t
ORIGIN      387 bp upstream of HindIII site on chromosome 12.

  DB 121;  Score      43;  Match 69.0%;  QryMatch 1.0%;  Pred. No. 4.78e-20;
  Matches    87;  Conservative    0;  Mismatches  38;  Indels   1;  Gaps   1;

Db   2558 gagagggagagggagagggagagggagagggagagggagagggagagggagagggagagg 2617
          || |||||  | ||  | || ||||| || ||  | || |||||  |||| |||||  |
Qy   3963 GATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGGGATCGT 4022

Db   2618 gagagggagagggagggggagagagagagagagagagagagagagagagagagagagaga 2677
          ||  |||| |||||  |||| |||||  ||  ||||||  |||| || | ||||  || |
Qy   4023 GACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG-GCGAGACCGATA 4081

Db   2678 gagaga 2683
           |||||
Qy   4082 TAGAGA 4087


RESULT   12
LOCUS       XLU170K0     1332 bp    DNA             VRT       05-APR-1990
DEFINITION  Xenopus laevis U1 70K gene exon 10.
ACCESSION   X12429
KEYWORDS    U1 small nuclear RNA.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
            Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amphibia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1332)
  AUTHORS   Etzerodt,M., Vignali,R., Scherly,D., Mattaj,I.W., Ciliberto,G. and
            Philipson,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1988) to the EMBL/GenBank/DDBJ databases.
            Etzerodt M., University of Aarhus, Dept. of Molecular Biology, C.F.
            Moellers Allee 130, DK-8000 Aarhus C, Denmark
REFERENCE   2  (bases 1 to 1332)
  AUTHORS   Etzerodt,M., Vignali,R., Ciliberto,G., Scherly,D., Mattaj,I.W. and
            Philipson,L.
  TITLE     Structure and expression of a Xenopus gene encoding an snRNP
            protein (U1 70K)
  JOURNAL   EMBO J. 7 (13), 4311-4321 (1988)
  MEDLINE   89210819
COMMENT     NCBI gi: 65172
FEATURES             Location/Qualifiers
     source          1..1332
                     /organism="Xenopus laevis"
                     /clone_lib="lambda EMBL4A"
```

```
                    /clone="Xg47"
     intron         <1..101
                    /note="intron IX"
     mRNA           102..1167
                    /note="exon 10"
     misc_feature   1145..1150
                    /note="poly A signal"
     polyA_site     1167
                    /note="poly A site"
BASE COUNT      436 a    235 c    352 g    309 t
ORIGIN

  DB 136;  Score      41;  Match 65.4%;  QryMatch 0.9%;  Pred. No. 8.13e-18;
  Matches    87;  Conservative    0;  Mismatches  46;  Indels   0;  Gaps   0;

Db    557 gaaggcgcagccatagagacagggatagagagaaggatagggatagggacagggacaggg 616
          |||||    || ||  ||| ||||||| ||||     || ||||| || ||  | |||||||
Qy   3946 GAAGGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGG 4005

Db    617 acaggaggcgagatagggaccgcgatagagagagggacaaggatcacaaacgagagcgag 676
          |  ||    | ||| | ||||| || || || ||||| |  || |      |||| ||||
Qy   4006 ATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAG 4065

Db    677 acagaggtgaccg 689
          ||||  | |||||
Qy   4066 ACAGGCGAGACCG 4078


RESULT  13
LOCUS       HUMUT1265     593 bp    DNA             STS       27-MAY-1993
DEFINITION  Human chromosome 2 STS UT1265.
ACCESSION   L16379
KEYWORDS    PCR primer; STS sequence; microsatellite marker;
            microsatellite repeat; repeat polymorphism; sequence tagged site;
            tetranucleotide repeat.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,
            Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsner,T.,
            Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.
  TITLE     Genetic and physical mapping of simple sequence repeat containing
            sequence tagged sites from the human genome
  JOURNAL   Unpublished (1993) See COMMENT for author address
COMMENT     Submitted by: Utah Center for Human Genome Research University of
            Utah, Dept. of Human Genetics
            2160 Eccles Institute of Human Genetics
            Salt Lake City, UT  84112
            e-mail: sts@corona.med.utah.edu
            Primer A: AATCACATGCGTCAACTCCT
            Primer B: GAGCCCAGGGTGAAAGAAG
            32P-label: B Primer
            PCR Profile:
            Initial Denaturation: 94C 300sec
            PCR Cycles: 5
            Denaturation: 94C  10sec
            Annealing: 60C  10sec
            Extension: 72C  20sec
            Mg++: 2mM
            Gel: Acrylamide 7%, Formamide 32%, Urea 34%
            Alleles: 6.
```



```
            NCBI gi: 307615
FEATURES            Location/Qualifiers
     source         1..593
                    /organism="Homo sapiens"
                    /sequenced_mol="DNA"
     STS            21..313
                    /standard_name="STS UT1265"
                    /map="2"
     primer_bind    21..40
     primer_bind    complement(295..313)
BASE COUNT      207 a    126 c    131 g    120 t      9 others
ORIGIN

  DB 76;  Score      41;  Match 68.1%;  QryMatch 0.9%;  Pred. No. 8.13e-18;
  Matches    77;  Conservative    0;  Mismatches  36;  Indels   0;  Gaps   0;

Db    115 gatagatatagagacagagacagagacagacatagaggcagagacagagatagagataga 174
          |||  | |||| ||  ||||| | || ||  | ||||   | ||||| ||| | |||||
Qy   3957 GATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGG 4016

Db    175 gatagagatagagatagagatagagatagagatagagctagagctagagatag 227
          ||| | ||  | || || ||||| |||||||||  || | ||||   |||| ||
Qy   4017 GATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG 4069


RESULT  14
LOCUS       MUSMHRDDK    1812 bp    DNA             ROD       16-DEC-1992
DEFINITION  Mouse MHC class III RD gene (H2-d and H2-Sk haplotypes), complete
            cds.
ACCESSION   M21332
KEYWORDS    RD protein; class III gene; major histocompatibility complex.
SOURCE      Mouse H2-d haplotype, DNA, clone WL10S and mouse H2-Sk haplotype
            liver, cDNA to mRNA, clone WL623.
  ORGANISM  Mus musculus
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE   1  (bases 1 to 1812)
  AUTHORS   Levi-Strauss,M., Carroll,M.C., Steinmetz,M. and Meo,T.
  TITLE     A previously undetected MHC gene with an unusual periodic structure
  JOURNAL   Science 240, 201-204 (1988)
  MEDLINE   88178091
COMMENT     Authors named the 42 kd polypeptide product RD for the most common
            dipeptide repeat.  Both haplotypes had identical sequences for the
            exons.

            NCBI gi: 199607
FEATURES            Location/Qualifiers
     source         1..1812
                    /organism="Mus musculus"
                    /haplotype="H2-d and H2-Sk"
                    /sequenced_mol="DNA"
                    /tissue_type="liver"
     exon           <1..131
                    /gene="RD"
                    /number=1
     intron         132..451
                    /gene="RD"
                    /number=1
     CDS            460..1587
                    /note="42 kd polypeptide (RD),  (first expressed exon);
                    NCBI gi: 199608"
                    /codon_start=1
                    /translation="MLVIPPGLSEEEEALQKEFNKLKKKKKALLALKKQSSSGPASQG
                    GVKRSLSEQPVVDTATATEQAKQLVKSGAISAIKAETKNSGFKRSRTLEGKLKDPEKG
```

```
                    PVPTFQPFQRSMSADEDLQEPSRRPQRKSLYESFVSSSDRLRELGQDGEEAEAPGAGD
                    GPPRGFDWSYEEHGSARSSASPPRSRSRDRSHDRSRDRDRDKERDRDRDRDRDRDRDK
                    DKDRDRDRDRDKERDRDRDRDRDRERDREGPFRRSDSFPERRAPRKGNTLYVYGEDMT
                    PTLLRGAFSPFGNIIDLSMDPPRNCAFVTYEKMESADQAVAELNGTQVESVQLKVNIA
                    RKQPMLDAATGKSVWGSLAVQNSPKGCHRDKRTQIVYSDDL"
BASE COUNT      445 a    478 c    542 g    347 t
ORIGIN      506 bp upstream of PstI site; chromosome 17.

  DB 121;  Score      41;  Match 65.4%;  QryMatch 0.9%;  Pred. No. 8.13e-18;
  Matches    87;  Conservative    0;  Mismatches  46;  Indels   0;  Gaps   0;

Db   1034 atgaccggagccgagatagggatcgggacaaagagagagacagagaccgagatcgagatc 1093
          |||| ||  |  | ||| | |||   || |  ||  ||||| | ||  | ||  ||||||
Qy   3937 ATGATCGTCGAAGGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATC 3996

Db   1094 gagacagagaccgggataaagacaaagacagagaccgggatcgagacagagacaaagagc 1153
          | ||||| || ||||||| ||    ||| | ||| ||||| | || |||||   |  |
Qy   3997 GCGACAGGGATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGA 4056

Db   1154 gggatcgagacag 1166
          | || ||||||||
Qy   4057 GAGAACGAGACAG 4069


RESULT   15
LOCUS       T02581        243 bp    mRNA            EST       26-MAY-1992
DEFINITION  0224C3 Plasmodium falciparum cDNA clone PF0224C.
ACCESSION   T02581
KEYWORDS    EST.
SOURCE      Malaria, falciparum clone=PF0224C library=cbsPfHB3.1, Debopam
            Chakrabarti vector=pBlueScript SK(+) host=E. coli XL1-Blue
            primer=T3 PolyA+ RNA, from asynchronous blood stage parasites of
            the cloned Honduran HB3 isolate cultured in vitro, was reverse
            transcribed using an oligo dT-Xho I primer. Second strand was
            prepared using RNase H and DNA polymerase I.  EcoR I adapters were
            ligated to the cDNA ,and it was cleaved with Xho I.  Prepared
            fragments were ligated into EcoR I + Xho I double-digested
            pBlueScript SK(+), and transformed E. coli XL1-Blue.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Animalia; Protozoa; Apicomplexa; Sporozoa; Coccidia;
            Eucoccidiida; Haemosporina; Plasmodiidae.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Dame,J.B.
  JOURNAL   Unpublished (1992)
COMMENT
            Contact: Dame JB
            Department of Infectious Diseases
            University of Florida
            Bldg. 471 Mowry Rd., Gainesville, FL  32611-0880
            Tel: 9043924700
            Fax: 9043929704
            Email: dame@icbr.ifas.ufl.edu.

            NCBI gi: 271435
FEATURES             Location/Qualifiers
     source          1..243
                     /organism="Plasmodium falciparum"
                     /clone="PF0224C"
                     /note="Malaria, falciparum"
BASE COUNT      113 a     18 c     80 g     29 t      3 others
ORIGIN

  DB 49;  Score      40;  Match 68.9%;  QryMatch 0.9%;  Pred. No. 1.02e-16;
  Matches    71;  Conservative    0;  Mismatches  32;  Indels   0;  Gaps   0;
```



```
Db    133 aagagatcgcaatagagatagagaacatagaagagatagggatagagatagagaaagaga 192
          ||| |||||| ||  |||||| || |      | || ||||| |||||| | || || ||
Qy   3947 AAGGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGA 4006

Db    193 tcgtaatagggacagngaaagtgatcgtaatagagatagagaa 235
          ||| ||||||| | ||  | ||  |  |||| ||||||||||
Qy   4007 TCGGGATAGGGATCGTGACCGGGACAGGGATAGGGATAGAGAA 4049

Search completed: Wed Jan 17 18:08:31 1996
Job time : 3512 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:22:34 1996; MasPar time 17.16 Seconds
502.247 Million cell updates/sec
Tabular output not generated.

Title: >US-08-319-745-10
Description: (1:1356) from US08319745.pep
Perfect Score: 9913
Sequence: 1 MASAGNARRGPGQAGRRREA.....TQRPPPWAALCPATASPSPL 1356

Scoring table: PAM 150
Gap 11

Searched: 53402 seqs, 6354270 residues

Database: a-geneseq18
1 part1
2 part2
3 part3
4 part4
5 part5
6 part6
7 part7
8 part8
9 part9
10 part10

Statistics: Mean 40.815; Variance 192.679; scale 0.212

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	115	1.2	439	5	R28150	Sugar beet chitinase	4.21e+00
2	114	1.2	1229	10	R54074	CryET5.	4.88e+00
3	114	1.2	928	1	R06289	Predicted retinoblast	4.88e+00
4	111	1.1	1280	2	P70452	Sequence encoded by h	7.60e+00
5	111	1.1	1280	8	R44297	Sequence encoded by h	7.60e+00
6	110	1.1	1280	1	R04868	Protein encoded by Mu	8.80e+00
7	109	1.1	521	8	R47068	Mammalian chromaffin	1.02e+01
8	108	1.1	487	4	R22380	Antigen mc-35c.	1.18e+01

9	108	1.1	195	8	R47339	Peptide fragment of m	1.18e+01
10	108	1.1	380	8	R48063	Sequence of polypepti	1.18e+01
11	106	1.1	928	1	R05305	Cancer supressing gen	1.57e+01
12	106	1.1	286	1	P82590	Polypeptide with glyc	1.57e+01
13	106	1.1	816	1	P82112	Human retinoblastoma	1.57e+01
14	106	1.1	928	7	R36534	Retinoblastoma (RB) p	1.57e+01
15	106	1.1	970	1	P90599	Human retinoblastoma.	1.57e+01
16	104	1.0	272	1	P93560	Plasmodium berghei ci	2.09e+01
17	103	1.0	380	3	R14857	Cry1X protein.	2.41e+01
18	101	1.0	458	10	R54834	Human derived adrenal	3.20e+01
19	101	1.0	253	2	R10535	Prod. of pMG3C9 used	3.20e+01
20	101	1.0	864	4	R24042	Lipoxygenase.	3.20e+01
21	100	1.0	719	2	R08041	81 kD endotoxin deduc	3.68e+01
22	99	1.0	402	2	P70644	Pseudorabies virus gp	4.24e+01
23	98	1.0	575	10	R57139	Interleukin-10 recept	4.87e+01
24	98	1.0	504	10	R54682	Mouse brain 5HT2C ser	4.87e+01
25	98	1.0	1253	5	R28337	SFV4 structural polyp	4.87e+01
26	98	1.0	685	2	R11331	Human luteinising hor	4.87e+01
27	97	1.0	316	3	R15878	Cholesterol esterase	5.59e+01
28	97	1.0	316	3	R15868	Cholesterol esterase	5.59e+01
29	97	1.0	316	3	R15873	Cholesterol esterase	5.59e+01
30	97	1.0	657	5	R28964	Notch hN5k full lengt	5.59e+01
31	97	1.0	316	3	R15866	Cholesterol esterase	5.59e+01
32	97	1.0	316	3	R15881	Cholesterol esterase	5.59e+01
33	97	1.0	316	3	R15870	Cholesterol esterase	5.59e+01
34	97	1.0	316	3	R15880	Cholesterol esterase	5.59e+01
35	97	1.0	316	3	R15883	Cholesterol esterase	5.59e+01
36	97	1.0	316	3	R15877	Cholesterol esterase	5.59e+01
37	97	1.0	316	3	R15869	Cholesterol esterase	5.59e+01
38	97	1.0	316	3	R15874	Cholesterol esterase	5.59e+01
39	97	1.0	316	3	R15865	Cholesterol esterase	5.59e+01
40	97	1.0	316	3	R15867	Cholesterol esterase	5.59e+01
41	97	1.0	316	3	R15882	Cholesterol esterase	5.59e+01
42	97	1.0	316	3	R15879	Cholesterol esterase	5.59e+01
43	97	1.0	316	3	R15872	Cholesterol esterase	5.59e+01
44	97	1.0	316	3	R15675	Cholesterol esterase	5.59e+01
45	97	1.0	457	3	P50361	Human acetyl choline	5.59e+01

ALIGNMENTS

```
RESULT    1
ID   R28150 standard; Protein; 439 AA.
AC   R28150;
DT   17-MAR-1993  (first entry)
DE   Sugar beet chitinase 1.
KW   SBC-1; fungicide; anti-fungal agent; extensine.
OS   Beta vulgaris cv monova.
FH   Key             Location/Qualifiers
FT   Peptide         1..26
FT   /label= leader
FT   Domain          27..46
FT   /label= hevein_domain
FT   Domain          47..178
FT   /label= proline_rich
FT   /note= "possibly involved in anchoring
FT   chitinase 1 protein to the cell wall
FT   after modification of the prolines to
FT   glycosylated hydroxyprolines, as in
FT   extensines"
FT   Domain          179..416
FT   /label= functional_domain
FT   Region          417..439
FT   /note= "probably directs the protein to
```

FT the vacuole"
PN WO9217591-A.
PD 15-OCT-1992.
PF 07-APR-1992; DK0108.
PR 08-APR-1991; DK-000616.
PA (DANI-) DANISCO AS.
PI Berglund L, Bojsen K, Mikkelsen JD, Nielsen KK;
DR WPI; 92-366261/44.
DR N-PSDB; Q29965.
PT Sugar beet chitinase 4 and corresponding DNA - inhibits
PT germination and growth of chitin-containing fungi and is used to
PT produce genetically transformed plants
PS Claim 5; Page 164-168; 254pp; English.
CC A clone encoding the chitinase 1 gene was isolated from a
CC sugar beet EMBL3 genomic library. The sequence encodes a protein
CC having 439 amino acid residues. Transgenic plants having increased
CC resistance to nematodes and chitin-containing plant pathogens,
CC partic. fungi, can be produced using genetic constructs containing
CC the chitinase 1 gene. The protein itself can be used in fungicidal
CC compositions.
SQ Sequence 439 AA;

DB 5; Score 115; Match 29.6%; QryMatch 1.2%; Pred. No. 4.21e+00;
Matches 29; Conservative 16; Mismatches 50; Indels 3; Gaps 3;

Db 67 ptprpppprpptprpppptprpppprp-ptprpppppptp-rpppprpptprpppppptprp 124
Qy 1258 PDSRHQPPLTPRQQPHLDSGSLSPGRQGQQPRRDPPREGLRPPPYRPRRDAFEISTEGHS 1317

Db 125 pppptprpppspp-tprpppppppspptpsppsppsp 161
Qy 1318 GPSNRDRSGPVGPVLTTLGTQRPPPWAALCPATASPSP 1355

RESULT 2
ID R54074 standard; Protein; 1229 AA.
AC R54074;
DT 02-FEB-1995 (first entry)
DE CryET5.
KW cryET4; cryET5; Lepidoptera; lepidopteran insect; insecticidal;
KW toxin; insecticidal crystal protein; ICP.
OS Bacillus thuringiensis.
PN US5322687-A.
PD 21-JUN-1994.
PF 29-JUL-1993; 100709.
PR 29-JUL-1993; US-100709.
PA (ECOG-) ECOGEN INC.
PI Donovan WP, Gonzalez JM, Jany CS, Tan Y;
DR WPI; 94-199503/24.
DR N-PSDB; Q64112.
PT Isolated cryET4 gene and Bacillus thuringiensis cultures
PT transformed with this gene - used in compsns. against
PT lepidopteran insects.
PS Disclosure; Col 29-38; 51pp; English.
CC B. thuringiensis strain EG5847 exhibits insecticidal activity
CC against lepidopteran insects. Two novel toxin genes from B.
CC thuringiensis EG5847 designated cryET4 and cryET5 produce
CC insecticidal proteins with activity against a broad spectrum of
CC lepidopteran insects. The gene sequences are given in Q64111-12.
SQ Sequence 1229 AA;

DB 10; Score 114; Match 24.4%; QryMatch 1.2%; Pred. No. 4.88e+00;
Matches 31; Conservative 38; Mismatches 49; Indels 9; Gaps 9;

Db 233 qwyntglnnlrgtnaes-wlrynqfrrdltlgvldlvalfpsydtrtypintsaqltrei 291
Qy 836 DWLQ-GLQDAFDSDWETGRIMPNNYKNGSDDGVLAYKLLVQT-GSRDKPIDIS-QLTKQR 892

Db 292 ytdpigrtnapsgfas-tnwfnnnapsfsaieaaifrpphlldfpeqltiysassr-w-s 348
Qy 893 LVDADGIINPSAFYIYLTAWVSNDPVAYAASQANI-RP-HRPEWVHDKADYMPETRLRIP 950

Db 349 stqhmny 355
Qy 951 AAEPIEY 957

RESULT 3
ID R06289 standard; protein; 928 AA.
AC R06289;
DT 13-DEC-1990 (first entry)
DE Predicted retinoblastoma gene product.
KW Osteosarcoma; fibrosarcoma; glioblastoma; breast cancer; ds;
OS Homo sapiens.
PN US4942123-A.
PD 17-JUL-1990.
PF 17-SEP-1987; 098612.
PR 17-SEP-1987; US-098612.
PA (REGC) UNIV OF CALIFORNIA.
PI Lee WH, Eva Y, Lee HP;
DR WPI; 90-245977/32.
PT Diagnosing absence or inactivation of retinoblastoma gene - by
PT detecting the absence of specific anti-ppRB 110 antibody
PT immuno-complex formed using tissue
PS Disclosure; p; English.
CC Laelled Abs raised to the RB gene product may be used to screen
CC for RB and in diagnosis of susceptibility to associated secondary
CC cancers such as osteosarcoma, fibrosarcoma, glioblastoma and
CC breast cancer.
SQ Sequence 928 AA;

DB 1; Score 114; Match 22.1%; QryMatch 1.2%; Pred. No. 4.88e+00;
Matches 31; Conservative 41; Mismatches 58; Indels 10; Gaps 9;

Db 669 llsenpelemiiwt-lfqhtlqneyewlmrdrhldqimcsmygickvknidlkfkiiv-t 726
Qy 819 MLEENKQLPQM-WLHYFRDWLQGLQDAFDSDWETGRIMPNNYKNGSDDGV-LAYKLLVQT 876

Db 727 aykdlphavqetfkrvlikeeeydsiivfynsvfm-qrlktnilqyastrpptlspiph- 784
Qy 877 GSRDKPIDISQLTKQRLVDADGIINPSAFY--IYLTAWVSNDPVAYAASQANIRPHRPEW 934

Db 785 ip-rspykfpssplripggn 803
Qy 935 VHDKADYM-PETRLRIPAAE 953

RESULT 4
ID P70452 standard; Protein; 1280 AA.
AC P70452;
DT 21-MAY-1991 (first entry)
DE Sequence encoded by human multi-drug resistance-1 (mdr1) cDNA
DE from clones lambda-HDR10,5 and 104.
KW Chemo-therapy resistant tumour cell; P-glycoprotein.
OS Homo sapiens.
PN WO8705943-A.
PD 08-OCT-1987.
PF 26-MAR-1987; U00758.

```
PR   28-MAR-1986; US-845610.
PR   01-AUG-1986; US-892575.
PA   (UNII ) UNIV OF ILLINOIS.
PI   Roninson IB, Pastan IH, Gottesman MM;
DR   WPI; 87-291656/41.
DR   N-PSDB; N70752.
PT   DNA for multi-drug resistance in human cells - used to detect
PT   chemotherapy-resistant tumour cells and for producing
PT   polypeptide(s) for diagnosis and therapy
PS   Claim 4(a); Table 5, pp30-39; 61pp; English.
CC   The human multi-drug resistant KB carcinoma cell lines were used as
CC   the source of the mdr1 gene nucleic acid sequences (N70751). To
CC   obtain cDNA clones of the mdr1 gene (N70752), poly (A) and RNA was
CC   used. Analysis of the AA sequence presented in (N70752) indicates
CC   that the mdr1 gene product is likely to be a transmembrane protein.
CC   The presence of transmembrane domains and potential glycosylation
CC   sites is consistent with the mdr1 protein being related to the
CC   P-glycoprotein.
SQ   Sequence   1280 AA;

  DB  2;  Score     111;  Match 22.8%;  QryMatch 1.1%;  Pred. No. 7.60e+00;
  Matches    29;  Conservative   43;  Mismatches  46;  Indels   9;  Gaps   8;

Db    216 lvilaispvlglsaavwakilssftdkellayakagavaeevlaairtviafggqkkele 275
          || |:::: ||| : : :  ::: :  ::|::   |   ::|:   :: ::  ||:| :
Qy    462 LVALSVAAGLGLCSLI-GISFNA-ATTQVLPFLALGVGVDDVFLLAHA-FSETGQNKRI- 517

Db    276 rynknlee-akrigikkaitanisigaaflliyasy-al-afwygttlvlsgeysigqvl 332
          ::    |  || |   |:|:  :: | |:    :  || ||   :::|:   :::: ||
Qy    518 PFEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVV--FNFAMVL 575

Db    333 tvffsvl 339
          :| ::|
Qy    576 LIFPAIL 582


RESULT    5
ID   R44297 standard; Protein; 1280 AA.
AC   R44297;
DT   24-JUN-1994  (first entry)
DE   Sequence encoded by human multi-drug resistant gene mdr1.
KW   p-glycoprotein; multi-drug resistant gene; mdr1.
OS   Homo sapiens.
PN   WO9324613-A.
PD   09-DEC-1993.
PF   14-MAY-1993; U04707.
PR   22-MAY-1992; US-887712.
PA   (GENE-) GENETIC THERAPY INC.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Mcdonagh KT,  Nienhuis A,  Tolstoshev P;
DR   WPI; 93-405805/50.
DR   N-PSDB; Q52726.
PT   DNA or RNA sequence for human multi-drug resistant gene MDR1 -
PT   encodes P-glycoprotein with altered splice site, used in cancer
PT   therapy
PS   Example; Fig 4; 64pp; English.
CC   pMDR2000 contains an mdr1 cDNA sequence (Q52726) described in PCT
CC   application no. WO87/0943m wherein the first 282 bp of the 5' UTR
CC   and the last 23 bp of the 3' UTR of the cDNA sequence have been
CC   removed.
SQ   Sequence   1280 AA;

  DB  8;  Score     111;  Match 22.8%;  QryMatch 1.1%;  Pred. No. 7.60e+00;
  Matches    29;  Conservative   43;  Mismatches  46;  Indels   9;  Gaps   8;
```



```
Db    216 lvilaispvlglsaavwakilssftdkellayakagavaeevlaairtviafggqkkele 275
          || |:::: ||| : : :  ::: :  ::|::   |   ::|:   :: ::  ||:| :
Qy    462 LVALSVAAGLGLCSLI-GISFNA-ATTQVLPFLALGVGVDDVFLLAHA-FSETGQNKRI- 517

Db    276 rynknlee-akrigikkaitanisigaaflliyasy-al-afwygttlvlsgeysigqvl 332
          ::    |  || |   |:|:  :: | |:    :  || ||   :::|:   :::: ||
Qy    518 PFEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVV--FNFAMVL 575

Db    333 tvffsvl 339
          :| ::|
Qy    576 LIFPAIL 582


RESULT    6
ID   R04868 standard; protein; 1280 AA.
AC   R04868;
DT   01-OCT-1990  (first entry)
DE   Protein encoded by Multidrug Resistance A1 gene
KW   Multidrug Resistance A1 gene; drug resistance of cancer cells; ss.
OS   Homo sapiens.
PN   J02100680-A.
PD   12-APR-1990.
PF   05-OCT-1988; 251475.
PR   05-OCT-1988; JO-251475.
PA   (SUNR) Suntory Ltd.
DR   WPI; 90-159707/21.
DR   N-PSDB; Q04522.
PT   MDR related gene derived from human normal cells -
PT   codes specific amino acid sequence, used for diagnosing drug
PT   resistance of cancer cells
PS   Disclosure; p; Japanese.
CC   The gene that encodes this protein is useful for diagnosis of drug
CC   resistance of cancer cells.
SQ   Sequence   1280 AA;

  DB  1;  Score     110;  Match 23.8%;  QryMatch 1.1%;  Pred. No. 8.80e+00;
  Matches    30;  Conservative   43;  Mismatches  46;  Indels   7;  Gaps   6;

Db    216 lvilaispvlglsaavwakilssftdeellayakagavaeevlaairtviafggqkkele 275
          || |:::: ||| : : :  ::: | : :|::   |   ::|:   :: ::  ||:| :
Qy    462 LVALSVAAGLGLCSLI-GISFNAATTQ-VLPFLALGVGVDDVFLLAHA-FSETGQNKRIP 518

Db    276 rynksleeakrigikkaitanisigaaflliyasy-al-afwygttlvlsgeysigqvlt 333
          :::  |  || |   |:|:  :: | |:    :  || ||   :::|:   :::: ||
Qy    519 FEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVV--FNFAMVLL 576

Db    334 vffsvl 339
          :| ::|
Qy    577 IFPAIL 582


RESULT    7
ID   R47068 standard; Protein; 521 AA.
AC   R47068;
DT   01-JUL-1994  (first entry)
DE   Mammalian chromaffin granule amine transporter protein.
KW   Vesicle membrane transport protein; gene therapy; screening;
KW   Parkinsons disease; neurotoxin; identification; detection;
KW   antibody; probe;chromaffin granule amine transporter protein.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   Domain          22..42
FT   /label= Transmembrane domain.
```

```
FT   Modified_site   58
FT   /note= "Site of N-linked glycosylation."
FT   Modified_site   87
FT   /note= "Site of N-linked glycosylation."
FT   Modified_site   104
FT   /note= "Site of N-linked glycosylation."
FT   Domain          136..153
FT   /label= Transmembrane domain.
FT   Domain          165..185
FT   /label= Transmembrane domain.
FT   Domain          195..215
FT   /label= Transmembrane domain.
FT   Domain          225..247
FT   /label= Transmembrane domain.
FT   Domain          254..276
FT   /label= Transmembrane domain.
FT   Domain          297..316
FT   /label= Transmembrane domain.
FT   Domain          333..357
FT   /label= Transmembrane domain.
FT   Domain          363..383
FT   /label= Transmembrane domain.
FT   Domain          395..415
FT   /label= Transmembrane domain.
FT   Domain          420..440
FT   /label= Transmembrane domain.
FT   Domain          446..467
PN   WO9325699-A.
PD   23-DEC-1993.
PF   11-JUN-1993; U05704.
PR   11-JUN-1992; US-899074.
PR   30-JUL-1992; US-923096.
PA   (REGC ) UNIV CALIFORNIA.
PI   Edwards RH;
DR   WPI; 94-007556/01.
DR   N-PSDB; Q54621.
PT   New mammalian vesicle membrane transport protein - and corresp.
PT   DNA, vectors, transformed cells and antibodies, for diagnosis and
PT   treatment of neurological disorders, e.g. Parkinson disease
PS   Claim 2; Figure 1; 181pp; English.
CC   The cDNA encoding the protein is useful in gene therapy and as a
CC   probe for detecting genomic sequences.  The protein is used
CC   for screening cytotoxic compounds implicated in Parkinsons disease,
CC   diseases associated with activity of neurotoxins or psychiatric
CC   disorders and to identify compounds which selectively inhibit or
CC   activate its action.  Antibodies raised against this protein are
CC   useful as immunoassay reagents for detecting the protein and as
CC   affinity reagents for purification.
SQ   Sequence   521 AA;

  DB  8;  Score     109;  Match 26.0%;  QryMatch 1.1%;  Pred. No. 1.02e+01;
  Matches    19;  Conservative   23;  Mismatches  26;  Indels   5;  Gaps   5;

Db    370 mvavgisllcvplahnifgligp-nagl-gfaigmvdsslmpimgylvdlrhtsvygsvy 427
          :::::: | |  |   :| |: |  ||:    ::::   |  :|| |: :: ::|   |
Qy   1015 LLSISVVLACTFLVCAVF-LLNPWTAGIIVMVLALMTVELFGMMG-LIGIKLSAV-PVVI 1071

Db    428 aiadvafcvgfai 440
           || |:: | |::
Qy   1072 LIASVGIGVEFTV 1084


RESULT    8
ID   R22380 standard; Protein; 487 AA.
```



```
AC   R22380;
DT   17-AUG-1992  (first entry)
DE   Antigen mc-35c.
KW   Oocysts; sporozoite; beta-galactosidase.
OS   Eimeria maxima.
FH   Key             Location/Qualifiers
FT   Peptide         2..9
FT   /label= repeat_unit
FT   Region          2..193
FT   /label= repeat_region
FT   Peptide         195..202
FT   /label= repeat_unit
FT   Region          195..322
FT   /label= repeat_region
PN   WO9204461-A.
PD   19-MAR-1992.
PF   05-SEP-1991; U06431.
PR   12-SEP-1990; US-581694.
PA   (GENE-) GENEX CORP.
PA   (FARH ) HOECHST AG.
PI   Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI   Raether W;
DR   WPI; 92-114366/14.
DR   N-PSDB; Q23080.
PT   Vaccine against avian coccidiosis - comprising recombinant
PT   Eimeria antigen mc-4c, mc-5c or mc-30c gene, etc., or
PT   microorganisms expressing them
PS   Claim 8; Page 64 + Fig 8; 94pp; English.
CC   To identify antigens of E. maxima, expression libraries were prepd.
CC   in lambda vector, lambda gt11, using cDNA prepd. from polyA mRNA
CC   isolated from E. maxima oocysts.  The cDNA expression library was
CC   screened with rat antiserum raised against the sporozoite stage of
CC   E. maxima.  The library to be screened was plated on a host that
CC   allows lysis and plaque formation.  Following induction of the
CC   antigens encoded by the phage, the plaques were transferred to
CC   nitrocellulose filters.  Positive phage were identified after
CC   screening with the rat anti-E. maxima sporozoite antiserum.
CC   The cDNA inserts from the positive clones were cloned into
CC   bacteriophage M13 and subjected to sequence analysis.
CC   E. maxima antigen mc-35c was identified.
SQ   Sequence   487 AA;

  DB  4;  Score     108;  Match 23.3%;  QryMatch 1.1%;  Pred. No. 1.18e+01;
  Matches    31;  Conservative   34;  Mismatches  65;  Indels   3;  Gaps   2;

Db    234 ppqrpghgppqrpghgppqrpghgppqrpghgppqrpghgppqrpghgppqrpghgppqr 293
          || :  :|:    :  ::|    | ::   :  :|: : ||:::    : : |  : :
Qy   1196 PPGHTNNGSDSSDSEYSSQTTVSGISEELRQYEAQQGAGGPAHQVIVEATENPVFARSTV 1255

Db    294 pghgppqrpghgppqrpg-hgppqrpghdgghpsrgsgrgglipkrfagrpdrgseqnqe 352
             : ::|   | |:|   : :  ||::| :| | : | || |  :  || | : :
Qy   1256 VHPDSRHQPPLTPRQQPHLDSGSLSPGRQGQQPRRDPPREGLRPPPY--RPRRDAFEIST 1313

Db    353 eeqsgggqnstrsa 365
          | :||  |  ||:
Qy   1314 EGHSGPSNRDRSG 1326


RESULT    9
ID   R47339 standard; Protein; 195 AA.
AC   R47339;
DT   01-JUL-1994  (first entry)
DE   Peptide fragment of multi-drug resistance (BMR) transporter protein.
KW   Vesicle membrane transport protein; gene therapy; screening;
```

```
KW   Parkinsons disease; neurotoxin; identification; detection;
KW   antibody; probe;chromaffin granule amine transporter protein;
KW   resistance; multi-drug resistance; BMR.
OS   Bacillus subtilis.
PN   WO9325699-A.
PD   23-DEC-1993.
PF   11-JUN-1993; U05704.
PR   11-JUN-1992; US-899074.
PR   30-JUL-1992; US-923096.
PA   (REGC ) UNIV CALIFORNIA.
PI   Edwards RH;
DR   WPI; 94-007556/01.
PT   New mammalian vesicle membrane transport protein - and corresp.
PT   DNA, vectors, transformed cells and antibodies, for diagnosis and
PT   treatment of neurological disorders, e.g. Parkinson disease
PS   Example 2; Page 113-114; 181pp; English.
CC   This sequence of the bacterial multi-drug resistance transporter
CC   protein showed a definite homology with the chromaffin granule
CC   amine transport protein (CGAT) of rat. The cDNA encoding the
CC   chromaffin granule amine transport protein is useful in gene therapy
CC   and as a probe for detecting genomic sequences. CGAT is used for
CC   screening cytotoxic compounds implicated in Parkinsons disease,
CC   diseases associated with activity of neurotoxins or psychiatric
CC   disorders and to identify compounds which selectively inhibit or
CC   activate its action. Antibodies raised against the transport
CC   protein are useful as immunoassay reagents for detecting the protein
CC   and as affinity reagents for purification.
SQ   Sequence  195 AA;

  DB  8;  Score     108;  Match 23.9%;  QryMatch 1.1%;  Pred. No. 1.18e+01;
  Matches    28;  Conservative   36;  Mismatches  46;  Indels   7;  Gaps   7;

Db     64 wvdrfgrkimivigllffsvseflfgig-ktvemlfitrmlggisapfimpgvtafiadi 122
          |   :  ::: :: : :|::   |:||   :| ::::   :| |:  | :    ||:: |
Qy   1037 WTAGIIVMVLALMTVELFGMMG-LIGIKLSAVPVVILIASVG-IGVEFTVHVALAFLTAI 1094

Db    123 ttiktrpk-algymsaaistgfiigpgiggflaevhsrlpfffaaafallaailsil 178
             : |:: ||  | |::  |  ::  :| ::    | : |:    ||:|| ||::|
Qy   1095 GDKNHRAMLALEHMFAPVLDG-AVSTLLGVLMLAG-SEFDFIVRYFFAVLA-ILTVL 1148


RESULT   10
ID   R48063 standard; Protein; 380 AA.
AC   R48063;
DT   20-JUL-1994  (first entry)
DE   Sequence of polypeptide encoded by the first open reading frame in
DE   the unique short (Us) region of bovine herpes virus (BHV) genome
KW   Insertion region; unique short region; Us; vaccine; antigen.
OS   Bovine herpes virus type 1, strain ST.
PN   WO9400586-A.
PD   06-JAN-1994.
PF   25-JUN-1993; F00642.
PR   26-JUN-1992; FR-007930.
PA   (INMR ) RHONE MERIEUX SA.
PI   Audonnet JF,  Legastelois ICMA,  Leung-tack P,  Riviere MEA;
DR   WPI; 94-026222/03.
DR   N-PSDB; Q53350.
PT   New insertion region sequence of bovine herpes virus genomic DNA
PT   - used for recombinant virus with this region deleted or
PT   inactivated, useful in vaccines allowing differentiation between
PT   vaccinated and infected cattle
PS   Claim 9; Fig 2; 47pp; French.
CC   The 4190 bp sequence in Q55350 encodes polypeptides homologous to
CC   HSV-1 gI, gE and US9; these are neither essential for in vitro
```



```
CC   replication nor important in inducing a protective immune response.
CC   It is genomic DNA purified from the ST strain of BHV-1. (BHV is also
CC   known as infectious bovine rhinotracheitis virus.). The sequence
CC   forms the insertion region in genomic DNA. rBHV in which the
CC   specific insertion region, esp. nucleotides 172-1311, has been
CC   deleted or inactivated by insertion are claimed.
SQ   Sequence  380 AA;

  DB  8;  Score     108;  Match 25.9%;  QryMatch 1.1%;  Pred. No. 1.18e+01;
  Matches    30;  Conservative   32;  Mismatches  47;  Indels   7;  Gaps   7;

Db    172 epgdpeaaartpapsrqsrpaasgltssasly-dr-alarspqappprpappraa-ragp 228
          |: :    ||:     :||  :: ||:      |  :|: : |:  ||  ||| : |: |
Qy   1243 EATENPVFARSTVVHPDSRHQPP-LTPRQQPHLDSGSLSPGRQGQQPRRDPPREGLRPPP 1301

Db    229 rrpervdetteveaatragsafalttppagpt-aspaaspsrafsaaapaaaaqpa 283
           || | | : |: :  ::|::    : | ||  :: ::  : : :|  ||:|: ::
Qy   1302 YRPRR-D-AFEISTEGHSGPSNRDRSGPVGPVLTTLGTQRPPPWAALCPATASPSP 1355


RESULT   11
ID   R05305 standard; protein; 928 AA.
AC   R05305;
DT   11-OCT-1990  (first entry)
DE   Cancer supressing gene (CSG) product.
KW   Cancer; cancer supressing gene; CSG; 13q14; retinoblastoma;
KW   RB; ds.
OS   Homo sapiens.
PN   WO9005180-A.
PD   17-MAY-1990.
PF   30-OCT-1989; 004808.
PR   31-OCT-1988; US-265829.
PA   (REGC) Univ of Caifornia.
PI   Lee WH, Huang HJS;
DR   WPI; 90-178822/23.
DR   N-PSDB; Q04713.
PT   Controlling cancer -
PT   by replacing ineffective cancer suppressing gene with cloned,
PT   active gene.
PS   Claim 35; Page 86; 105pp; English.
CC   Gene is taken from human chromosome 13q14 retinoblastoma (RB) cDNA.
CC   By installing a working CSG, safe and specific treatment and
CC   prophylaxis can be given to cancer patients.
SQ   Sequence  928 AA;

  DB  1;  Score     106;  Match 21.3%;  QryMatch 1.1%;  Pred. No. 1.57e+01;
  Matches    30;  Conservative   43;  Mismatches  56;  Indels  12;  Gaps  11;

Db    669 llsehpelehiiwt-lfqhtlqn-eyelmrdrhldqimmcsmygickvknidlkfkiiv- 725
          :| |: :| :: |   |:  ||  :  :  |:   :||  : |       : | :|::|
Qy    819 MLEENKQLPQM-WLHYFRDWLQGLQDAFDSDWETGRIMP-NNYKNGSDDGV-LAYKLLVQ 875

Db    726 taykdlphavqetfkrvlikeeeydsiivfynsvfm-qrlktnilqyastrpptlspiph 784
          |: :| |  : :  |: |:  :   :   ||  :::  ::  : : ||::::   :  |
Qy    876 TGSRDKPIDISQLTKQRLVDADGIINPSAFY--IYLTAWVSNDPVAYAASQANIRPHRPE 933

Db    785 -ip-rspykfpssplripggn 803
           :  :: |: | : ||||:::
Qy    934 WVHDKADYM-PETRLRIPAAE 953


RESULT   12
ID   P82590 standard; protein; 286 AA.
AC   P82590;
```

```
DT   05-NOV-1990  (first entry)
DE   Polypeptide with glycosylation inhibiting factor activity.
KW   Glycosylation; GIF; TGIF4; IgE; immunoglobulin; antiinflammatory agent.
PN   EP-285405-A.
PD   05-OCT-1988.
PF   30-MAR-1988; 302862.
PR   31-MAR-1987; US-032859.
PA   (UYJO) John Hopkins University.
PI   Ishizaka K, Martens CL, Moore KW;
DR   WPI; 88-280165/40.
DR   N-PSDB; N81168.
PT   Nucleic acid encoding glycosylation inhibiting peptide - capable of
PT   suppressing IgE immunoglobulin responses, for use as antiinflammatory
PT   agents.
PS   Disclosure; ; 29pp; English.
CC   The polypeptide has the folllowing biological activities: IgE binding
CC   factor switching activity,inhibition of IgE-induced Fc-epsilon receptor
CC   expression, and inhibition of phospholipase A2 activity after
CC   dephosphorylation.
SQ   Sequence   286 AA;

  DB  1;  Score     106;  Match 32.1%;  QryMatch 1.1%;  Pred. No. 1.57e+01;
  Matches    17;  Conservative   13;  Mismatches  19;  Indels   4;  Gaps   4;

Db    168 ysyyglssipsmrpylwwkkyitqgqlvqfvltiiqttcgvfwpcs-fplg-w 218
          |:  |||| |:  |:| |  ||:  : : : :::: : |  |  |: | |  |
Qy    987 YTSLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLA-C-TFLVCAVFLLNPW 1037


RESULT   13
ID   P82112 standard; protein; 816 AA.
AC   P82112;
DT   22-OCT-1990  (first entry)
DE   Human retinoblastoma gene product
KW   Retinoblastoma; neoplasm; osteocarcoma.
OS   Homo sapiens.
PN   EP-259031-A.
PD   09-MAR-1988.
PF   11-AUG-1987; 307095.
PR   11-AUG-1986; US-895163.
PA   (MASS-) Massachusets Eye; (WHIT-) Whitehead Inst Biomed Re.
PI   Dryja TP, Friend S;
DR   WPI; 88-065827/10.
DR   N-PSDB; N81369.
PT   Diagnosis of retinoblastoma - using genetic material corresp. to a normal
PT   human retinoblastoma gene or a unique sub-region
PS   Disclosure; p; English.
CC   The absence of this protein is associated with a neoplasm. The corresp.
CC   gene is used to produce an antibody which is contacted with the tumour
CC   sample and immune complexes,if detected, indicate the presence in the
CC   tumour sample of the protein. The whole sequence or a unique sub-region
CC   can be used. Method identifies patients lacking the defective retino-
CC   blastoma (Rb) allele and thus are not at risk of developing the disease.
SQ   Sequence   816 AA;

  DB  1;  Score     106;  Match 21.3%;  QryMatch 1.1%;  Pred. No. 1.57e+01;
  Matches    30;  Conservative   43;  Mismatches  56;  Indels  12;  Gaps  11;

Db    557 llsehpelehiiwt-lfqhtlqn-eyelmrdrhldqimmcsmygickvknidlkfkiiv- 613
          :| |: :| :: |   |:  ||  :  :  |:   :||  : |       : | :|::|
Qy    819 MLEENKQLPQM-WLHYFRDWLQGLQDAFDSDWETGRIMP-NNYKNGSDDGV-LAYKLLVQ 875

Db    614 taykdlphavqetfkrvlikeeeydsiivfynsvfm-qrlktnilqyastrpptlspiph 672
          |: :| |  : :  |: |:  :   :   ||  :::  ::  : : ||::::   :  |
```



```
Qy    876 TGSRDKPIDISQLTKQRLVDADGIINPSAFY--IYLTAWVSNDPVAYAASQANIRPHRPE 933

Db    673 -ip-rspykfpssplripggn 691
           :  :: |: | : ||||:::
Qy    934 WVHDKADYM-PETRLRIPAAE 953


RESULT   14
ID   R36534 standard; Protein; 928 AA.
AC   R36534;
DT   24-AUG-1993  (first entry)
DE   Retinoblastoma (RB) protein.
KW   RB gene product; p56RB portion; cell cycle progression control;
KW   combination; therapeutic methods; arrest; tumourigenesis;
KW   regulation; physiological processes; blood cell prodn.;
KW   gamete prodn.; ss.
OS   Homo sapiens SR-40 cell line.
PN   WO9308267-A.
PD   29-APR-1993.
PF   16-OCT-1992; U08918.
PR   17-OCT-1991; US-778510.
PA   (REGC ) UNIV CALIFORNIA.
PI   Goodrich DW, Lee EYHP, Lee WH, Wang NP.
DR   WPI; 93-152462/18.
DR   N-PSDB; Q41545.
PT   Method of controlling cell cycle progression - uses purified
PT   retinoblastoma protein or fragment, for use in combination with
PT   therapeutic methods to arrest tumorigenesis
PS   Claim 10; Fig 9; 68pp; English.
CC   The sequence is that of the retinoblastoma (RB) protein which may be
CC   used as part of a method of controlling cell cycle progression which
CC   may be used in combination with therapeutic methods to arrest
CC   tumourigenesis in organisms. The cell cycle can be reversibly
CC   arrested in a convenient and safe manner. The protein is used in
CC   a compsn. is relatively inexpensive and readily obtainable, and
CC   shows little or no toxic effects on healthy cells. It is also
CC   compatible with other methods and devices for regulating certain
CC   physiological processes of the body, such as blood cell prodn. and
CC   gamete prodn. Fragments of the protein are soluble in low concns.
CC   of glycerol thus enhancing their value in pharmaceutical applicns.
SQ   Sequence   928 AA;

  DB  7;  Score     106;  Match 21.3%;  QryMatch 1.1%;  Pred. No. 1.57e+01;
  Matches    30;  Conservative   43;  Mismatches  56;  Indels  12;  Gaps  11;

Db    669 llsehpelehiiwt-lfqhtlqn-eyelmrdrhldqimmcsmygickvknidlkfkiiv- 725
          :| |: :| :: |   |:  ||  :  :  |:   :||  : |       : | :|::|
Qy    819 MLEENKQLPQM-WLHYFRDWLQGLQDAFDSDWETGRIMP-NNYKNGSDDGV-LAYKLLVQ 875

Db    726 taykdlphavqetfkrvlikeeeydsiivfynsvfm-qrlktnilqyastrpptlspiph 784
          |: :| |  : :  |: |:  :   :   ||  :::  ::  : : ||::::   :  |
Qy    876 TGSRDKPIDISQLTKQRLVDADGIINPSAFY--IYLTAWVSNDPVAYAASQANIRPHRPE 933

Db    785 -ip-rspykfpssplripggn 803
           :  :: |: | : ||||:::
Qy    934 WVHDKADYM-PETRLRIPAAE 953


RESULT   15
ID   P90599 standard; protein; 970 AA.
AC   P90599;
DT   30-OCT-1989  (first entry)
DE   Human retinoblastoma.
KW   Human retinoblastoma; screening; tumours; probes.
```

OS Homo sapiens (human).
PN WO8906703-A.
PD 27-JUL-1989.
PF 23-JAN-1989; U00293.
PR 21-JAN-1988; US-146525.
PA (DRYJ) Dryja T P.
PI Dryja T P; Friend S; Yandell D W.
DR WPI; 89-233856/32.
DR N-PSDB; N90489, N90490.
PT New human retino-blastoma gene and polypeptide(s)
PT - used for screening individuals for defective retino-blastoma
PT gene and treating these patients.
PS Disclosure; fig 5; 71 pp; English.
CC Human retinoblastoma (RB) (see N90489 and N90490).
CC This can be used to screen individuals for the presence of
CC the mutated RB gene. The RB polypeptide can prevent retinoblastoma
CC formation, and corresp. antibodies can be used in tumour immunodiagnosis.
SQ Sequence 970 AA;

DB 1; Score 106; Match 21.3%; QryMatch 1.1%; Pred. No. 1.57e+01;
Matches 30; Conservative 43; Mismatches 56; Indels 12; Gaps 11;

```
Db    711 llsehpelehiiwt-lfqhtlqn-eyelmrdrhldqimmcsmygickvknidlkfkiiv- 767
          :| |: :| :: |   |:  || :  :  |:   :||  : |       : | :|::|
Qy    819 MLEENKQLPQM-WLHYFRDWLQGLQDAFDSDWETGRIMP-NNYKNGSDDGV-LAYKLLVQ 875

Db    768 taykdlphavqetfkrvlikeeeydsiivfynsvfm-qrlktnilqyastrpptlspiph 826
          |: :| |  : :  |: |:  :   :   || :::  ::  : : ||::::   :  |
Qy    876 TGSRDKPIDISQLTKQRLVDADGIINPSAFY--IYLTAWVSNDPVAYAASQANIRPHRPE 933

Db    827 -ip-rspykfpssplripggn 845
           :  :: |: | : ||||:::
Qy    934 WVHDKADYM-PETRLRIPAAE 953
```

Search completed: Wed Jan 17 17:23:52 1996
Job time : 78 secs.

****

MPSRCH (TM)

****

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:17:06 1996; MasPar time 28.56 Seconds
728.090 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-10
Description: (1:1356) from US08319745.pep
Perfect Score: 9913
Sequence: 1 MASAGNARRGPGQAGRRREA.........TQRPPPWAALCPATASPSPL 1356

Scoring table: PAM 150
Gap 11

Searched: 43470 seqs, 15335248 residues

Database: swiss-prot31
1 part1
2 part2
3 part3
4 part4
5 part5
6 part6
7 part7
8 part8

Statistics: Mean 57.851; Variance 125.966; scale 0.459

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	2806	28.3	1286	5	PATC_DROME	MEMBRANE PROTEIN PATC	0.00e+00
2	236	2.4	633	8	YLS3_CAEEL	HYPOTHETICAL 70.7 KD	5.26e-17
3	143	1.4	382	2	DACB_BACSU	PENICILLIN-BINDING PR	6.30e-04
4	137	1.4	712	2	CYAB_BORPE	CYCLOLYSIN SECRETION	3.40e-03
5	135	1.4	215	5	NOLH_RHIME	NODULATION PROTEIN NO	5.90e-03
6	134	1.4	413	8	YMB5_CAEEL	HYPOTHETICAL 47.0 KD	7.76e-03
7	130	1.3	472	1	ARAE_ECOLI	ARABINOSE-PROTON SYMP	2.29e-02
8	130	1.3	1063	2	CZCA_ALCEU	CATION EFFLUX SYSTEM	2.29e-02
9	128	1.3	549	8	YJCG_ECOLI	HYPOTHETICAL 59.2 KD	3.90e-02
10	128	1.3	1034	1	ACRF_ECOLI	ACRIFLAVIN RESISTANCE	3.90e-02
11	128	1.3	217	8	YKR4_EBV	HYPOTHETICAL BKRF4 PR	3.90e-02
12	127	1.3	286	2	CYSW_SYNP7	SULFATE TRANSPORT SYS	5.08e-02
13	126	1.3	445	8	YIEG_ECOLI	HYPOTHETICAL 46.9 KD	6.61e-02
14	123	1.2	246	1	ATP6_CANPA	ATP SYNTHASE A CHAIN	1.44e-01
15	121	1.2	435	8	YYBO_BACSU	HYPOTHETICAL 48.2 KD	2.40e-01
16	119	1.2	915	8	YLS4_CAEEL	HYPOTHETICAL 102.9 KD	3.99e-01
17	119	1.2	1595	7	SOS_DROME	SON OF SEVENLESS PROT	3.99e-01
18	117	1.2	666	5	MXIA_SHIFL	MXIA PROTEIN (VIRH PR	6.56e-01
19	117	1.2	464	3	GALP_ECOLI	GALACTOSE-PROTON SYMP	6.56e-01
20	117	1.2	1711	2	CHD1_MOUSE	CHROMODOMAIN-HELICASE	6.56e-01
21	117	1.2	199	5	NU6C_PLEBO	NADH-PLASTOQUINONE OX	6.56e-01
22	116	1.2	307	1	BRAD_PSEAE	HIGH-AFFINITY BRANCHE	8.40e-01
23	116	1.2	126	8	YIGF_ECOLI	HYPOTHETICAL 14.5 KD	8.40e-01
24	115	1.2	381	2	CYB_CYPCA	CYTOCHROME B (EC 1.10	1.07e+00
25	115	1.2	436	7	SECY_MICLU	PREPROTEIN TRANSLOCAS	1.07e+00
26	112	1.1	532	8	YABM_BACSU	HYPOTHETICAL 57.4 KD	2.21e+00
27	112	1.1	1063	6	POLS_RUBVR	STRUCTURAL POLYPROTEI	2.21e+00
28	112	1.1	1733	8	VNUA_PRVKA	PROBABLE NUCLEAR ANTI	2.21e+00
29	112	1.1	454	5	NOLI_RHIME	NODULATION PROTEIN NO	2.21e+00
30	112	1.1	254	3	GUFA_MYXXA	GUFA PROTEIN.	2.21e+00
31	111	1.1	349	8	YHHT_ECOLI	HYPOTHETICAL 38.5 KD	2.81e+00
32	111	1.1	514	7	VE2_HPV5B	E2 PROTEIN.	2.81e+00
33	111	1.1	1280	4	MDR1_HUMAN	MULTIDRUG RESISTANCE	2.81e+00
34	110	1.1	696	4	LSHR_PIG	LUTROPIN-CHORIOGONADO	3.56e+00
35	110	1.1	241	1	ATP6_RHORU	ATP SYNTHASE A CHAIN	3.56e+00
36	110	1.1	354	6	PROW_ECOLI	GLYCINE BETAINE/L-PRO	3.56e+00
37	110	1.1	187	3	HAPP_PHYPO	PLASMODIUM-SPECIFIC H	3.56e+00
38	110	1.1	448	3	GNTP_BACSU	GLUCONATE PERMEASE.	3.56e+00
39	109	1.1	1049	1	ACRB_ECOLI	ACRIFLAVIN RESISTANCE	4.49e+00
40	109	1.1	274	5	NU2M_DROMA	NADH-UBIQUINONE OXIDO	4.49e+00
41	109	1.1	521	8	VMT1_RAT	CHROMAFFIN GRANULE AM	4.49e+00
42	109	1.1	286	5	NU2M_DROME	NADH-UBIQUINONE OXIDO	4.49e+00
43	109	1.1	620	3	EXTN_TOBAC	EXTENSIN PRECURSOR (C	4.49e+00
44	109	1.1	1446	4	IE18_PRVKA	IMMEDIATE-EARLY PROTE	4.49e+00
45	109	1.1	253	8	YSO2_DESAM	HYPOTHETICAL 28.3 KD	4.49e+00

ALIGNMENTS

RESULT 1
ID PATC_DROME STANDARD; PRT; 1286 AA.
AC P18502;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE MEMBRANE PROTEIN PATCHED.
GN PTC.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RM 90058658
RA HOOPER J.E., SCOTT M.P.;
RL CELL 59:751-765(1989).
RN [2]
RP SEQUENCE FROM N.A.
RM 90015164
RA NAKANO Y., GUERRERO I., HIDALGO A., TAYLOR A., WHITTLE J.R.S.,
RA INGHAM P.W.;
RL NATURE 341:508-513(1989).
CC -!- FUNCTION: SEGMENTATION POLARITY PROTEIN. EXACT FUNCTIONNOT
CC KNOWN. PTC PROBABLY PARTICIPATES IN CELL INTERACTIONS THAT
CC ESTABLISH PATTERN WITHIN THE SEGMENT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

```
CC   -!- SIMILARITY: TO C.ELEGANS SEX-DETERMINING TRANSFORMER PROTEIN 2.
DR   EMBL; M28418; DMMPP1.
DR   EMBL; M28999; DMMPP2.
DR   EMBL; X17558; DMPTCR.
DR   PIR; A33468; A33468.
DR   PIR; S06119; S06119.
DR   FLYBASE; FBGN0003892; TUF.
KW   TRANSMEMBRANE; DEVELOPMENTAL PROTEIN; SEGMENTATION POLARITY PROTEIN.
FT   TRANSMEM     77     92       POTENTIAL.
FT   TRANSMEM    428    448       POTENTIAL.
FT   TRANSMEM    466    486       POTENTIAL.
FT   TRANSMEM    493    513       POTENTIAL.
FT   TRANSMEM    533    553       POTENTIAL.
FT   TRANSMEM    563    583       POTENTIAL.
FT   TRANSMEM    678    699       POTENTIAL.
FT   TRANSMEM    932    952       POTENTIAL.
FT   TRANSMEM    956    998       POTENTIAL.
FT   TRANSMEM   1008   1028       POTENTIAL.
FT   TRANSMEM   1057   1077       POTENTIAL.
FT   TRANSMEM   1083   1103       POTENTIAL.
FT   CARBOHYD    142    142       POTENTIAL.
FT   CARBOHYD    298    298       POTENTIAL.
FT   CARBOHYD    335    335       POTENTIAL.
FT   CARBOHYD    388    388       POTENTIAL.
FT   CARBOHYD    807    807       POTENTIAL.
FT   CARBOHYD   1181   1181       POTENTIAL.
FT   CARBOHYD   1258   1258       POTENTIAL.
FT   CARBOHYD   1283   1283       POTENTIAL.
FT   CONFLICT    111    111       R -> G (IN REF. 2).
FT   CONFLICT    274    274       G -> A (IN REF. 2).
FT   CONFLICT    332    332       R -> A (IN REF. 2).
FT   CONFLICT    636    636       P -> A (IN REF. 2).
FT   CONFLICT    862    864       DVF -> ASSPTELLRANCIRNR (IN REF. 2).
FT   CONFLICT    866    866       Y -> N (IN REF. 2).
SQ   SEQUENCE   1286 AA;  142915 MW;  8981681 CN;
```

DB 5; Score 2806; Match 40.7%; QryMatch 28.3%; Pred. No. 0.00e+00;
Matches 475; Conservative 301; Mismatches 301; Indels 90; Gaps 53;

```
Db     37 aldqidkgkargsrtaiylrsvfqshletlgssvqkhagkvlfvailvlstfcvglksaq 96
Qy     47 ALEQISKGKATGRKAPLWLRAKFQRLLFKLGCYIQKNCGKFLVVGLLIFGAFAVGLKAAN 106

Db     97 ihskvhqlwiqeggrleaelaytqktigedesathqlliqtthdpnasvlhpqallahle 156
Qy    107 LETNVEELWVEVGGRVSRELNYTRQKIGEEAMFNPQLMIQTPKEEGANVLTTEALLQHLD 166

Db    157 vlvkatavkvhlydtewglrdmcnmpstpsfegiyyieqilrhlipcsiitpldcfwegs 216
Qy    167 SALQASRVHVYMYNRQWKLEHLCYKSGELITETGY-MDQIIEYLYPCLIITPLDCFWEGA 225

Db    217 qllgpesavvipglnqrllwttlnpasvmqymkqkmseekisfdfetveqymkraaigsg 276
Qy    226 KLQS-GTAYLL-G-KPPLRWTNFDP---LEFLEE-LK--KINYQVDSWEEMLNKAEVGHG 276

Db    277 ymekpclnplnpncpdtapnknstqppdvgailsggcygyaakhmhwpeelivggrkrnr 336
Qy    277 YMDRPCLNPADPDCPATAPNKNSTKPLDVALVLNGGCQGLSRKYMHWQEELIVGGTVKNA 336

Db    337 sghlrkaqalqsvvqlmtekemydqwqd-nykvhhlgwtqekaaevlnawqrnfsreveq 395
Qy    337 TGKLVSAHALQTMFQLMTPKQMYEHFRGYDY-VSHINWNEDRAAAILEAWQRTYV-EVVH 394

Db    396 llrkqsriatnydiyvfssaalddilakfshpsalsivigvavtvlyafc-tllrwrdpv 454
```



```
Qy    395 --QSVAPNSTQ-KVLPFTTTTLDDILKSFSDVSVIRVASGYLLMLAYA-CLTMLRW-DCS 449

Db    455 rgqssvgvagvllmcfstaaglglsallgivfnaastqvvpflalglgvdhifmltaaya 514
Qy    450 KSQGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFS 509

Db    515 es--nrr---eq-tklilkkvgpsilfsacstagsffaaafipvpalkvfclqaaivmcs 568
Qy    510 ETGQNKRIPFEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVVF 569

Db    569 nlaaallvfpamisldlrrrtagradifcc----c----f---p-vwke-qpkv--app- 612
Qy    570 NFAMVLLIFPAILSMDLYRREDRRLDIFCCLTSPCVSRVIQVEPQAYTEPHSNTRYSPPP 629

Db    613 --v-lplnnn-n-g-rga---r---hpks---cnn-n-r----v-plp-aqnpll-e--q 646
Qy    630 PYTSHSFAHETHITMQSTVQLRTEYDPHTHVYYTTAEPRSEISVQPVTVTQDNLSCQSPE 689

Db    647 -----r---adipgss-hsl--a-s-fslatfafqhytpflmrswvkfltvmgflaalis 693
Qy    690 STSSTRDLLSQFSDSSLHCLEPPCTKWTLSSFAEKHYAPFLLKPKAKVVVILLFLGLLGV 749

Db    694 slyastrlqdgldiidlvpkdsnehkfldaqtrlfgfysmyavtqgnfeyptqqqllrdy 753
Qy    750 SLYGTTRVRDGLDLTDIVPRETREYDFIAAQFKYFSFYNMYIVTQKA-DYPNIQHLLYDL 808

Db    754 hdsfvrvphviknd ngglpdfwlllfsewlgnlqkifdeeyrdgrltkecwfpnassdai 813
Qy    809 HKSFSNVKYVMLEENKQLPQMWLHYFRDWLQGLQDAFDSDWETGRIMPNN-YKNGSDDGV 867

Db    814 layklivqtghvdnpvdkelvltnrlvnsdgiinqrafynylsawatndvfaygasqqgkl 873
Qy    868 LAYKLLVQTGSRDKPIDISQLTKQRLVDADGIINPSAFYIYLTAWVSNDPVAYAASQANI 927

Db    874 ypeprqyfhqpney--d--lkipkslplvyaqmpfylhgltdtsqiktlighirdlsvky 929
Qy    928 RPHRPEWVHDKADYMPETRLRIPAAEPIEYAQFPFYLNGLRDTSDFVEAIEKVRVICNNY 987

Db    930 egfglpnypsgipfifweqymtlrsslamilacvllaalvlvsllllsvwaavlvilsvl 989
Qy    988 TSLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAVFLLNPWTAGIIVM-VL 1046

Db    990 asla-qifgamtllgiklsaipavililsvgmmlcfnvlislgfmtsvgnrqrrvqlsmq 1048
Qy   1047 ALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVHVALAFLTAIGDKNHRAMLALE 1106

Db   1049 mslgplvhgmltsgvavfmlstspfefvirhfcwlllvvlcv-gacnsllvfpillsmvg 1107
Qy   1107 HMFAPVLDGAVSTLLGVLMLAGSEFDFIVRYF-FAVLAILTVLGVLNGLVLLPVLLSFFG 1165

Db   1108 peaelvplehpdristpsplpvrsskr 1134
Qy   1166 PCPEVSPANGLNRLPTPSPEPPPSVVR 1192


RESULT    2
ID   YLS3_CAEEL     STANDARD;      PRT;   633 AA.
AC   P34388;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 70.7 KD PROTEIN F09G8.3 IN CHROMOSOME III.
GN   F09G8.3.
```

```
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RM   94150718
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RL   NATURE 368:32-38(1994).
DR   EMBL; L11247; CEF09G8.
DR   PIR; S44795; S44795.
DR   WORMPEP; F09G8.3; CE00138.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   633 AA;  70660 MW;  2049957 CN;

  DB  8;  Score     236;  Match 27.8%;  QryMatch 2.4%;  Pred. No. 5.26e-17;
  Matches    50;  Conservative   49;  Mismatches  76;  Indels   5;  Gaps   4;

Db     58 gvkvyvystffpyyeqyltlsttvytlvvlvlfvafvtislflrvn-lagslvtvfvlls 116
          |:  |  :  | : |||::|   :   : :||  :|:  ::||     || :| |: |::
Qy    991 GLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAVFLLNPWTAGIIVMVLALMT 1050

Db    117 sylhlmewmyllgitvnvvsvinmamslgiaveffgqmlhgfynskkpkreerafaalvs 176
            |  |  | |:|| :: |:|: :  |:|||:|||  ::  :|  :   | : ||  ||
Qy   1051 VELFGM--MGLIGIKLSAVPVVILIASVGIGVEFTVHVALAFLTAIGDK-NHRAMLALEH 1107

Db    177 ngsttlsgifpaimitagclsfadsrvlityfcnqlvgiglvcavhgvvymptllaifgs 236
           :  | |  ::::     |: ::   :: ||   |  : ::  ::|:| :| ||::||:
Qy   1108 MFAPVLDGAVSTLLGVLM-LAGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLSFFGP 1166


RESULT    3
ID   DACB_BACSU     STANDARD;      PRT;   382 AA.
AC   P35150;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE
DE   CARBOXYPEPTIDASE) (EC 3.4.16.4) (DD-PEPTIDASE) (DD-CARBOXYPEPTIDASE)
DE   (PBP-5*).
GN   DACB.
OS   BACILLUS SUBTILIS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RM   92193254
RA   BUCHANAN C.E., LING M.-L.;
RL   J. BACTERIOL. 174:1717-1725(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / MARBURG;
RM   95020538
RA   SOROKIN A.V., ZUMSTEIN E., AZEVEDO V., EHRLICH S.D., SERROR P.;
RL   MOL. MICROBIOL. 10:385-395(1993).
CC   -!- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE
```



```
CC       CELL WALL PRECURSORS. REQUIRED SPECIFICALLY FOR THE SYNTHESIS OF
CC       THE SPORE FORM OF PEPTIDOGLYCAN (CORTEX).
CC   -!- CATALYTIC ACTIVITY: D-ALANYL-D-ALANINE + H(2)O = 2 D-ALANINE.
CC   -!- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. OUTER FORESPORE
CC       MEMBRANE.
CC   -!- TISSUE SPECIFICITY: LIKELY TO BE EXPRESSED SPECIFICALLY IN THE
CC       MOTHER CELL.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT ABOUT STAGE III OF SPORULATION.
CC   -!- SIMILARITY: WITH OTHER D-ALANYL-D-ALANINE CARBOXYPEPTIDASES.
DR   EMBL; L09228; BSDIA.
DR   EMBL; M84227; BSDACB.
DR   PIR; S45552; S45552.
DR   SUBTILIST; BG10527; DACB.
KW   HYDROLASE; CARBOXYPEPTIDASE; PEPTIDOGLYCAN SYNTHESIS; CELL WALL;
KW   SIGNAL; SPORULATION.
FT   SIGNAL        1     27
FT   CHAIN        28    382       PENICILLIN-BINDING PROTEIN 5*.
FT   ACT_SITE     60     60       ACYLATED BY PENICILLIN (BY SIMILARITY).
SQ   SEQUENCE   382 AA;  43081 MW;  775191 CN;

  DB  2;  Score     143;  Match 28.0%;  QryMatch 1.4%;  Pred. No. 6.30e-04;
  Matches    21;  Conservative   22;  Mismatches  28;  Indels   4;  Gaps   3;

Db    233 staskdgidliavtindpndwddhmkmfnyvfehyqtyliakkgdipklkgtfye-skaf 291
          :|  :||:||  :   :  : |     |:| |  |  |::: |:| |::   :|:  |:|
Qy    754 TTRVRDGLDLTDIVPRETREYDFIAAQFKY-FSFYNMYIVTQKADYPNIQHLLYDLHKSF 812

Db    292 ikrdityllteeeke 306
              :: |:: ||:|:
Qy    813 --SNVKYVMLEENKQ 825


RESULT    4
ID   CYAB_BORPE     STANDARD;      PRT;   712 AA.
AC   P18770;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   CYCLOLYSIN SECRETION ATP-BINDING PROTEIN CYAB.
GN   CYAB.
OS   BORDETELLA PERTUSSIS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   ALCALIGENACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=18323;
RM   89091151
RA   GLASER P., SAKAMOTO H., BELLALOU J., ULLMANN A., DANCHIN A.;
RL   EMBO J. 7:3997-4004(1988).
CC   -!- FUNCTION: INVOLVED IN THE EXPORT OF CALMODULIN-SENSITIVE
CC       ADENYLATE CYCLASE-HAEMOLYSIN (CYCLOLYSIN).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). BELONGS TO THE HLYB SUBFAMILY.
DR   PIR; S02386; BVBRCB.
DR   PROSITE; PS00211; ABC_TRANSPORTER.
KW   HEMOLYSIS; ADENYLATE CYCLASE; TRANSPORT; ATP-BINDING; TRANSMEMBRANE.
FT   NP_BIND     505    512       ATP (BY SIMILARITY).
SQ   SEQUENCE   712 AA;  77969 MW;  2375035 CN;

  DB  2;  Score     137;  Match 27.6%;  QryMatch 1.4%;  Pred. No. 3.40e-03;
  Matches    43;  Conservative   39;  Mismatches  62;  Indels  12;  Gaps  10;
```

```
Db    166 lvlqfislltplffqvvmdkvlvnnametlnvitvgflaailfeal-ltgirtylfahts 224
          |:| :  :|:  |: |     |:|     : |: :::::  ||  : | ||:  | |
Qy   1014 LLLSISVVLACTFL-VC-AVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIK--LSAVPV 1069

Db    225 skldvelgarlyahllrlplayfqarrvgdsvarvr-elehirafltgnavtvlldv-vf 282
            |   :|  :    ::::||:: |  :||   |    |||: | :   ||: || | ::
Qy   1070 VILIASVGIGV-EFTVHVALAFLTA--IGDKNHRAMLALEHMFAPVLDGAVSTLLGVLML 1126

Db    283 svv-f-iavmffysvkltlvvlaalpcyfllslvlt 316
          :   | : | :|::|   | ||: |    ||:::|:
Qy   1127 AGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLS 1162


RESULT    5
ID   NOLH_RHIME     STANDARD;      PRT;   215 AA.
AC   P25198;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE   NODULATION PROTEIN NOLH PRECURSOR.
GN   NOLH.
OS   RHIZOBIUM MELILOTI.
OG   PLASMID SYM PRME41B.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   RHIZOBIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK 631;
RM   91360053
RA   BAEV N., ENDRE G., PETROVICS G., BANFALVI Z., KONDOROSI A.;
RL   MOL. GEN. GENET. 228:113-124(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRODUCTION OF MEDICAGO-SPECIFIC
CC       NODULATION SIGNAL MOLECULE.
DR   EMBL; X58632; RMPRME41B.
DR   PIR; S16564; S16564.
KW   PLASMID; NODULATION; SIGNAL.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    215       NODULATION PROTEIN NOLH.
SQ   SEQUENCE   215 AA;  23775 MW;  262628 CN;

  DB  5;  Score     135;  Match 23.4%;  QryMatch 1.4%;  Pred. No. 5.90e-03;
  Matches    32;  Conservative   47;  Mismatches  49;  Indels   9;  Gaps   8;

Db      1 mfl-nswrstvitgltlpisvigtfaaiyalgftlnimtlmalslsigiliddtivvren 59
          :|| |:| : :|  : |::  :  |: :  :|: |: : :: |  |:|| :: |: |
Qy   1031 VFLLNPWTAGIIV-MVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVHVALA 1089

Db     60 itrhlqmgkdpvraaldgtneiglavlsttlc-ivavflpvafmggligrfflqfgvtva 118
          :   :   |:  || |   : :: :||  ::  :::|::  :    :| |:|  |:| :|
Qy   1090 FLTAIG-DKNH-RAMLALEHMFA-PVLDGAVSTLLGVLMLAGSEFDFIVRYF--FAV-LA 1143

Db    119 vavvislfvsftldpml 135
          :  |:::: :: | |:|
Qy   1144 ILTVLGVLNGLVLLPVL 1160


RESULT    6
ID   YMB5_CAEEL     STANDARD;      PRT;   413 AA.
AC   Q03602;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 47.0 KD PROTEIN F54G8.5 IN CHROMOSOME III.
GN   F54G8.5.
```



```
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RM   94150718
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RL   NATURE 368:32-38(1994).
DR   EMBL; Z19155; CEF54G8.
DR   PIR; S28276; S28276.
DR   WORMPEP; F54G8.5; CE00207.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   413 AA;  46966 MW;  921564 CN;

  DB  8;  Score     134;  Match 29.6%;  QryMatch 1.4%;  Pred. No. 7.76e-03;
  Matches    24;  Conservative   24;  Mismatches  31;  Indels   2;  Gaps   2;

Db      9 eqryihaltesaaslfltsltdglsfaigsisdfhavrvfctycamailfmflfqvtffn 68
          |:|    |  ::|||: |||:::  :| ::::  : |:| |    |: ::| |   : :|
Qy    520 EDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVVFNFAMVLLIFP 579

Db     69 avmslc-crrevsgkhpvfcc 88
          |::|:   |||   :  :|||
Qy    580 AILSMDLYRREDR-RLDIFCC 599


RESULT    7
ID   ARAE_ECOLI     STANDARD;      PRT;   472 AA.
AC   P09830;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   ARABINOSE-PROTON SYMPORT (ARABINOSE TRANSPORTER).
GN   ARAE.
OS   ESCHERICHIA COLI.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   87115869
RA   MAIDEN M.C.J., DAVIS E.O., BALDWIN S.A., MOORE D.C.M.,
RA   HENDERSON P.J.F.;
RL   NATURE 325:641-643(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RM   88228015
RA   MAIDEN M.C.J., JONES-MORTIMER M.C., HENDERSON P.J.F.;
RL   J. BIOL. CHEM. 263:8003-8010(1988).
RN   [3]
RP   PRELIMINARY SEQUENCE OF 1-28 FROM N.A.
RM   84114868
RA   STONER C., SCHLEIF R.F.;
RL   J. MOL. BIOL. 171:369-381(1983).
```

```
CC   -!- FUNCTION: UPTAKE OF ARABINOSE ACROSS THE BOUNDARY MEMBRANE WITH
CC       THE CONCOMMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; J03732; ECARAEA.
DR   EMBL; X00272; ECARAE.
DR   PIR; B26430; B26430.
DR   PIR; A28075; A28075.
DR   ECOGENE; EG10056; ARAE.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2.
KW   TRANSPORT; SUGAR TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE; SYMPORT.
FT   TRANSMEM     23     43       POTENTIAL.
FT   TRANSMEM     63     83       POTENTIAL.
FT   TRANSMEM     92    110       POTENTIAL.
FT   TRANSMEM    120    138       POTENTIAL.
FT   TRANSMEM    149    167       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
FT   TRANSMEM    259    278       POTENTIAL.
FT   TRANSMEM    299    318       POTENTIAL.
FT   TRANSMEM    328    348       POTENTIAL.
FT   TRANSMEM    362    382       POTENTIAL.
FT   TRANSMEM    396    415       POTENTIAL.
FT   TRANSMEM    424    443       POTENTIAL.
SQ   SEQUENCE   472 AA;  51684 MW;  1186772 CN;

  DB  1;  Score     130;  Match 23.6%;  QryMatch 1.3%;  Pred. No. 2.29e-02;
  Matches    42;  Conservative   61;  Mismatches  56;  Indels  19;  Gaps  17;

Db     37 gldigviagalpfit-dhfv-ltsrlq-e-wvv-ssmmlgaaigalfngwlsfrlgrkys 91
          :| ::  : : ||:  :::: |  :|     || :   |  |:  |:| | :  :    :
Qy    989 SLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAV-FLLNPWTAGIIVMVLA 1047

Db     92 lmagailfvlgsigsafat-svemliaarvvlgiavgiasytaplylsemasenvrgkm- 149
          ||:  :: ::| ||  ::: :| :|||:  | ||:| ::   |  :|: ::  | |:::
Qy   1048 LMTVELFGMMGLIGIKLSAVPVVILIAS--V-GIGVEFTVHVALAFLTAIGDKNHRAMLA 1104

Db    150 is-my-qlm-vtlgivla-f-lsdtafsysgnwr-amlgvlalpavl--liilvvflp 199
          :  |:  ::  ::: :|: : |: : | :      |:|::|:: :||  |::| |:|:
Qy   1105 LEHMFAPVLDGAVSTLLGVLMLAGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLS 1162


RESULT    8
ID   CZCA_ALCEU     STANDARD;      PRT;  1063 AA.
AC   P13511;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CATION EFFLUX SYSTEM PROTEIN CZCA.
GN   CZCA.
OS   ALCALIGENES EUTROPHUS.
OG   PLASMID PMOL30.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   ALCALIGENACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CH34;
RM   90017477
RA   NIES D.H., NIES A., CHU L., SILVER S.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:7351-7355(1989).
CC   -!- FUNCTION: CZCA HAS LOW CATION TRANSPORT ACTIVITY FOR CO(2+), IT
CC       IS ESSENTIAL FOR THE EXPRESSION OF COBALT, ZINC, AND CADMIUM
CC       RESISTANCE. CZCA AND CZCB TOGETHER WOULD ACT IN ZN(2+) EFFLUX
CC       NEARLY AS EFFECTIVELY AS THE COMPLETE CZC EFFLUX SYSTEM (CZCABC).
```



```
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: HIGH, TO A.EUTROPHUS NICKEL AND COBALT RESISTANCE
CC       PROTEIN CNRA.
DR   EMBL; M26073; AECZC.
DR   PIR; A33830; A33830.
KW   PLASMID; TRANSPORT; ZINC; CADMIUM; COBALT; TRANSMEMBRANE.
FT   TRANSMEM     14     29       POTENTIAL.
FT   TRANSMEM    350    387       POTENTIAL.
FT   TRANSMEM    476    508       POTENTIAL.
FT   TRANSMEM    534    559       POTENTIAL.
FT   TRANSMEM    883    900       POTENTIAL.
FT   TRANSMEM    906    957       POTENTIAL.
FT   TRANSMEM    982   1004       POTENTIAL.
SQ   SEQUENCE  1063 AA;  115644 MW;  5749672 CN;

  DB  2;  Score     130;  Match 25.6%;  QryMatch 1.3%;  Pred. No. 2.29e-02;
  Matches    23;  Conservative   30;  Mismatches  34;  Indels   3;  Gaps   3;

Db    457 liimivylpifaltgvegkmfhpmaftvvlallgamilsvtfvpaaval-figervaeke 515
          :|:|:: |    | |: | :   ::   |: |:::: ::| |    ||| |:   :::|:
Qy   1041 IIVMVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFT-VHVALAFLTA-IGDKN 1098

Db    516 nrlmlwakrryeplleklslantavvltfaa 545
          :| ||   : : |:|: :::    || :|:
Qy   1099 HRAMLALEHMFAPVLDGAVSTLLGVLMLAG 1128


RESULT    9
ID   YJCG_ECOLI     STANDARD;      PRT;   549 AA.
AC   P32705;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 59.2 KD PROTEIN IN SOXR-ACS INTERGENIC REGION (F549).
GN   YJCG.
OS   ESCHERICHIA COLI.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RM   94089392
RA   BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,
RA   DANIELS D.L.;
RL   NUCLEIC ACIDS RES. 21:5408-5417(1993).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
DR   EMBL; U00006; ECUW89.
DR   ECOGENE; EG11942; YJCG.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2.
KW   HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE; SODIUM TRANSPORT;
KW   SYMPORT.
SQ   SEQUENCE   549 AA;  59197 MW;  1560344 CN;

  DB  8;  Score     128;  Match 28.1%;  QryMatch 1.3%;  Pred. No. 3.90e-02;
  Matches    34;  Conservative   37;  Mismatches  40;  Indels  10;  Gaps   9;

Db    172 liellfglnyhiavvlvgvlmmmyv-lfggm-lattwvqiikavlllfgasf-mafmvmk 228
          |:  :| ||   | ::| || :| | ||| | |    :  :  |:|: : :: : | |
Qy   1027 LVCAVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTV-- 1084

Db    229 hvgfsfnnlfseamavhpk-gvdimkpgglvkdpisalslglglmfgtaglphilmrfft 287
          ||:::|   ::: :  ::: ::: | :  |   ::|:| ||: ||:: : :  |:  ||:
```

```
Qy   1085 HVALAFLTAIGD-KNHRAMLALEHMFAPVL-DGAVSTL-LGV-LMLAGSEFDFIVRYFFA 1140

Db    288 v 288
          |
Qy   1141 V 1141


RESULT   10
ID   ACRF_ECOLI     STANDARD;      PRT;  1034 AA.
AC   P24181;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   ACRIFLAVIN RESISTANCE PROTEIN F (ENVD PROTEIN).
GN   ACRF OR ENVD.
OS   ESCHERICHIA COLI.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   XU J., NILLES M.L., BERTRAND K.P.;
RL   SUBMITTED (XXX-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=K12;
RM   92079901
RA   KLEIN J.R., HENRICH B., PLAPP R.;
RL   MOL. GEN. GENET. 230:230-240(1991).
RN   [3]
RP   REVISIONS.
RM   94012493
RA   MA D., COOK D.N., ALBERTI M., PON N.G., NIKAIDO H., HEARST J.E.;
RL   J. BACTERIOL. 175:6299-6313(1993).
CC   -!- FUNCTION: INVOLVED IN CELL ENVELOPE FORMATION. IS PRODUCED IN
CC       EXTREMELY LOW AMOUNTS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC       CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC       FRAMESHIFTS.
DR   EMBL; M96848; ECACREF.
DR   EMBL; X57948; ECENVCD.
DR   PIR; S18537; S18537.
DR   ECOGENE; EG10267; ACRF.
KW   CELL DIVISION; TRANSMEMBRANE; INNER MEMBRANE; TRANSPORT.
SQ   SEQUENCE   1034 AA;  111454 MW;  5626423 CN;

  DB  1;  Score     128;  Match 24.1%;  QryMatch 1.3%;  Pred. No. 3.90e-02;
  Matches    28;  Conservative   38;  Mismatches  42;  Indels   8;  Gaps   8;

Db    876 vaisfvvvfl-claalyeswsipvsvmlvvplgivgvllaatlfnqkndvyfmvglltti 934
          ::: :   || |   | ::|:  : || |::|  | ::    |:  |     :| |::::
Qy   1018 ISVVLACTFLVCAVFLLNPWTAGIIVM-VLALMTVELFGMMGLIGIKLSAVPVVILIASV 1076

Db    935 glsaknaili-vefakdlmekegkgvveatlmavrmrlrpilmtslafilgvlpla 989
          |::   :: : : |   : :|: :::: | |   :| : |:|  ::: :|||| ||
Qy   1077 GIGVEFTVHVALAFLTAIGDKNHRAML-A-L-E-HM-FAPVLDGAVSTLLGVLMLA 1127


RESULT   11
ID   YKR4_EBV       STANDARD;      PRT;  217 AA.
AC   P30117;
DT   01-APR-1993 (REL. 25, CREATED)
```



```
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL BKRF4 PROTEIN.
GN   BKRF4.
OS   EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   84270667
RA   BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA   GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA   TUFFNELL P.S., BARRELL B.G.;
RL   NATURE 310:207-211(1984).
CC   -!- SIMILARITY: TO HVS-1 GENE 45.
DR   EMBL; V01555; EBV.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   217 AA;  23980 MW;  256435 CN;

  DB  8;  Score     128;  Match 23.0%;  QryMatch 1.3%;  Pred. No. 3.90e-02;
  Matches    34;  Conservative   38;  Mismatches  70;  Indels   6;  Gaps   6;

Db     43 qeedvsdtdesdysdedeeidleeeyps-dedpsegsdsdpswhpsdsdesdysesdede 101
          : :: ||: :|:|| :     : ||    : : : |: :        :::: :: |
Qy   1199 HTNNGSDSSDSEYSSQTTVSGISEELRQYEAQQGAGGPAHQV-IVEATENPVFARSTVVH 1257

Db    102 atpgsqasrssrvspstqqssgltptpsfsrprtrapprppa-papvrgrasapprppap 160
          : :  |:: ::|  |  : |::|:|     :|| | |||    |:| | |  |   :
Qy   1258 PDSRHQPPLTPRQQPHLD-SGSLSPGRQGQQPR-RDPPREGLRPPPYRPRRDAFEISTEG 1315

Db    161 vqqstkdkgphrptrpvlrgpaprrppp 188
          : :  ::    |  |||   : :||||
Qy   1316 -HSGPSNRDRSGPVGPVLTTLGTQRPPP 1342


RESULT   12
ID   CYSW_SYNP7     STANDARD;      PRT;  286 AA.
AC   P27370;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   SULFATE TRANSPORT SYSTEM PERMEASE PROTEIN CYSW.
GN   CYSW.
OS   SYNECHOCOCCUS SP. (STRAIN PCC 7942) (ANACYSTIS NIDULANS R2).
OC   PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;
OC   CYANOBACTERIA (BLUE-GREEN ALGAE); CHROOCOCCALES.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   91210162
RA   LAUDENBACH D.E., GROSSMAN A.R.;
RL   J. BACTERIOL. 173:2739-2750(1991).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR SULFATE AND THIOSULFATE. PROBABLY RESPONSIBLE FOR THE
CC       TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- INDUCTION: BY SULFUR DEPRIVATION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC       CONTAINS 5 OR 6 POTENTIAL TRANSMEMBRANE DOMAINS (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYSTW/POTBC
CC       SUBFAMILY.
DR   EMBL; M65247; SSCYS.
DR   PIR; F43670; F43670.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR.
KW   INNER MEMBRANE; TRANSMEMBRANE; SULFATE TRANSPORT; TRANSPORT.
SQ   SEQUENCE   286 AA;  30671 MW;  448191 CN;
```

```
  DB  2;  Score     127;  Match 25.7%;  QryMatch 1.3%;  Pred. No. 5.08e-02;
  Matches    28;  Conservative   42;  Mismatches  31;  Indels   8;  Gaps   8;

Db     23 laligisllyvgliiiipaanvavqafsegls-gfiknlgdrnlqeairlt-llmgvisv 80
          ::|||| |  | ::|:|::  ::|: |:  :: :|:  :||:| :  : |  ::  |:
Qy   1057 MGLIGIKLSAVPVVILIASVGIGVE-FTVHVALAFLTAIGDKNHRAMLALEHMFAPVLDG 1115

Db     81 plntlfglaaafaiarkqfpgkslllsvidlpfsispvvagl-mi-vll 127
          :::||:|:   :| :  :|  :  :::|: : :::  |: || :: |||
Qy   1116 AVSTLLGVLM-LAGSEFDFIVR-YFFAVLAI-LTVLGVLNGLVLLPVLL 1161


RESULT   13
ID   YIEG_ECOLI     STANDARD;      PRT;   445 AA.
AC   P31466;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 46.9 KD PROTEIN IN TNAB-BGLB INTERGENIC REGION.
GN   YIEG.
OS   ESCHERICHIA COLI.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RM   93315143
RA   BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RL   GENOMICS 16:551-561(1993).
DR   EMBL; L10328; ECUW82.
DR   ECOGENE; EG11724; YIEG.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   445 AA;  46919 MW;  1139701 CN;

  DB  8;  Score     126;  Match 30.8%;  QryMatch 1.3%;  Pred. No. 6.61e-02;
  Matches    44;  Conservative   38;  Mismatches  50;  Indels  11;  Gaps  10;

Db    124 lllltifrvrywmianipvslrvgitsgiglfigmmglknagvivanpetlvsig-nlts 182
          :|: ::| :  |  : | : | :     :|: :|::|:| ::| |    : |:|| ::|
Qy   1026 FLVCAVFLLNPWTAGIIVMVLALMTVELFGM-MGLIGIKLSAVPVVILIASVGIGVEFTV 1084

Db    183 hsvl--lgilgf-fiiailasrnihaavlvsivvttllg-wmlgdvhyngivsappsvmt 238
          |  |  |  :|     |:||  :: |:|| :  |:|||  ||:   :: ||    :|::
Qy   1085 HVALAFLTAIGDKNHRAMLALEHMFAPVLDG-AVSTLLGVLMLAGSEFDFIVRYFFAVLA 1143

Db    239 vvghvdlagsfn-lglagvifsf 260
          ::  | | | :| | |  |::||
Qy   1144 ILT-V-L-GVLNGLVLLPVLLSF 1163


RESULT   14
ID   ATP6_CANPA     STANDARD;      PRT;   246 AA.
AC   Q03671;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   ATP SYNTHASE A CHAIN PRECURSOR (EC 3.6.1.34) (PROTEIN 6).
GN   ATP6 OR OLI2.
OS   CANDIDA PARAPSILOSIS (YEAST).
OG   MITOCHONDRION.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 4-25.
```



```
RC   STRAIN=CBS 7154 / SP1;
RM   91200034
RA   GUELIN E., GUERIN M., VELOURS J.;
RL   EUR. J. BIOCHEM. 197:105-111(1991).
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; X55653; CPATP6.
DR   PIR; S15378; S15378.
DR   PROSITE; PS00449; ATPASE_A.
KW   HYDROGEN ION TRANSPORT; CF(0); MITOCHONDRION; TRANSMEMBRANE.
FT   PROPEP        1      3       REMOVED IN MATURE FORM.
FT   CHAIN         4    246       ATP SYNTHASE A CHAIN.
FT   TRANSMEM    157    182       POTENTIAL.
FT   TRANSMEM    213    238       POTENTIAL.
SQ   SEQUENCE   246 AA;  26949 MW;  348051 CN;

  DB  1;  Score     123;  Match 24.2%;  QryMatch 1.2%;  Pred. No. 1.44e-01;
  Matches    31;  Conservative   39;  Mismatches  50;  Indels   8;  Gaps   8;

Db    112 lvavvsfsltlwi-gnvvlglylhgwgffalfvpsgtplalvpvlvliealsyasr-ais 169
          ||  | | |  |  | :|: | |    :|::    |  |: |||::|| ::: :   ::
Qy   1027 LVCAV-FLLNPWTAGIIVMVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVH 1085

Db    170 lglrlganils-ghllmlilgsliislmsssflgfvsgiipilavvait-ilefgiaiiq 227
          ::| : : |    |  || |  :: ::: ::   :: |:: :||   :  |: : :|::
Qy   1086 VALAFLTAIGDKNHRAMLALEHMFAPVLDGAVSTLL-GVL-MLAGSEFDFIVRYFFAVLA 1143

Db    228 ay-vfsil 234
            |:::|
Qy   1144 ILTVLGVL 1151


RESULT   15
ID   YYBO_BACSU     STANDARD;      PRT;   435 AA.
AC   P37489;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 48.2 KD PROTEIN IN COTF-TETB INTERGENIC REGION.
GN   YYBO.
OS   BACILLUS SUBTILIS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   OGASAWARA N., NAKAI S., YOSHIKAWA H.;
RL   DNA RES. 1:1-14(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
DR   EMBL; D26185; BSPURA1.
DR   SUBTILIST; BG10016; YYBO.
KW   HYPOTHETICAL PROTEIN; TRANSMEMBRANE; TRANSPORT.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     61     81       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    119    139       POTENTIAL.
FT   TRANSMEM    150    170       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
```

```
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    325    345       POTENTIAL.
FT   TRANSMEM    347    367       POTENTIAL.
FT   TRANSMEM    385    405       POTENTIAL.
FT   TRANSMEM    407    427       POTENTIAL.
SQ   SEQUENCE   435 AA;  48248 MW;  1105824 CN;

  DB  8;  Score     121;  Match 19.9%;  QryMatch 1.2%;  Pred. No. 2.40e-01;
  Matches    29;  Conservative   43;  Mismatches  68;  Indels   6;  Gaps   6;

Db    286 wlistisgiavggwlv-dyfikkgypntkvyrtviivgms-fgfffigsiltnnitvaii 343
          ||: :|| :     ||   |: :      :   : :: :  ||   | :|  : : | |:
Qy   1013 WLLLSISVVLACTFLVCAVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIKLSAVPVVIL 1072

Db    344 cisiglagisatapvgwsisaelapigsvsmlssmvnlannlfggiiaasltgylfdvtg 403
            |:|: |:  |  |: :: : ::     :||: : ::   :: | ::: |   ::  :
Qy   1073 IASVGI-GVEFTVHVALAFLTAIGDKNHRAMLA-LEHMFAPVLDGAVSTLLGVLMLAGSE 1130

Db    404 -sftlsflvagfvlllgl-vfyvfvl 427
           | : :: | : :|  | |:  :||
Qy   1131 FDFIVRYFFAVLAILTVLGVLNGLVL 1156

Search completed: Wed Jan 17 17:18:47 1996
Job time : 101 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:19:05 1996; MasPar time 45.33 Seconds
713.356 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-10
Description: (1:1356) from US08319745.pep
Perfect Score: 9913
Sequence: 1 MASAGNARRGPGQAGRRREA.....TQRPPPWAALCPATASPSPL 1356

Scoring table: PAM 150
Gap 11

Searched: 78488 seqs, 23849247 residues

Database: pir45
1 ann1
2 ann2
3 ann3
4 unann1
5 unann2
6 unann3
7 unann4
8 unann5
9 unann6
10 unann7
11 unrev1
12 unrev2

Statistics: Mean 55.272; Variance 153.526; scale 0.360

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	2806	28.3	1286	9	A33468	probable membrane pr	0.00e+00
2	2703	27.3	1299	9	S06119	membrane protein pat	0.00e+00
3	410	4.1	1170	11	S52525	hypothetical protein	7.72e-36
4	236	2.4	633	9	S44795	F09G8.3 protein - Ca	4.06e-13
5	145	1.5	464	8	S22697	extensin - Volvox ca	8.31e-03
6	143	1.4	382	7	S45552	penicillin-binding p	1.32e-02

7	143	1.4	382	7	B42274	penicillin-binding p	1.32e-02
8	137	1.4	711	8	C40046	antibiotic transport	5.23e-02
9	137	1.4	712	3	BVBRCB	cyaB protein - Borde	5.23e-02
10	135	1.4	215	7	S16564	nolH protein - Rhizo	8.20e-02
11	134	1.4	413	9	S28276	hypothetical protein	1.03e-01
12	130	1.3	472	7	B26430	arabinose transport	2.49e-01
13	130	1.3	403	12	S52796	prpL_2 protein - hum	2.49e-01
14	130	1.3	1063	7	A33830	cation efflux system	2.49e-01
15	129	1.3	295	10	B48013	proline-rich proteog	3.09e-01
16	128	1.3	226	6	S33024	hypothetical protein	3.85e-01
17	128	1.3	964	7	S18537	envD protein - Esche	3.85e-01
18	128	1.3	464	10	A47655	spliceosome-associat	3.85e-01
19	127	1.3	286	7	F43670	integral membrane pr	4.78e-01
20	123	1.2	1188	8	S49915	extensin-like protei	1.12e+00
21	123	1.2	246	11	S15378	H+-transporting ATP	1.12e+00
22	119	1.2	915	9	S44797	F09G8.4 protein - Ca	2.60e+00
23	119	1.2	353	9	S36438	EPPT protein - hydro	2.60e+00
24	119	1.2	1596	9	A41216	guanine nucleotide e	2.60e+00
25	118	1.2	430	8	JC2301	hypothetical 47.8K p	3.19e+00
26	117	1.2	199	7	JQ2137	NADH dehydrogenase (	3.92e+00
27	117	1.2	1711	10	A47392	chromodomain-helicas	3.92e+00
28	117	1.2	666	7	A44797	low-calcium-response	3.92e+00
29	116	1.2	502	10	A55197	Wiskott-Aldrich synd	4.81e+00
30	116	1.2	307	6	B36125	branched-chain amino	4.81e+00
31	116	1.2	126	7	S30744	hypothetical protein	4.81e+00
32	116	1.2	126	11	S30707	hypothetical protein	4.81e+00
33	115	1.2	439	8	S45025	chitinase (EC 3.2.1.	5.89e+00
34	115	1.2	381	1	S36011	ubiquinol--cytochrom	5.89e+00
35	115	1.2	436	6	S29891	secY protein - Micro	5.89e+00
36	114	1.2	226	7	C29504	hypothetical 24K pro	7.20e+00
37	112	1.1	454	7	S16565	nolI protein - Rhizo	1.07e+01
38	112	1.1	472	7	S47089	arabinose-proton sym	1.07e+01
39	112	1.1	254	7	S35738	gufA protein - Myxoc	1.07e+01
40	112	1.1	1063	3	GNWVRA	structural polyprote	1.07e+01
41	112	1.1	254	7	S39876	hypothetical protein	1.07e+01
42	112	1.1	1733	3	B45344	probable nuclear ant	1.07e+01
43	112	1.1	254	7	S33148	gufA protein - Myxoc	1.07e+01
44	111	1.1	376	7	S47693	hypothetical protein	1.31e+01
45	111	1.1	514	3	W2WLB5	E2 protein - human p	1.31e+01

ALIGNMENTS

RESULT 1
ENTRY A33468 #type complete
TITLE probable membrane protein patched - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 27-Jan-1995
ACCESSIONS A33468
REFERENCE A33468
#authors Hooper, J.E.; Scott, M.P.
#journal Cell (1989) 59:751-765
#title The Drosophila patched gene encodes a putative membrane protein required for segmental patterning.
#cross-references MUID:90058658
#accession A33468
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-1286 ##label HOO
##cross-references GB:M28418; GB:M28999
##note nucleotide sequence is not given
KEYWORDS membrane protein

```
SUMMARY          #length 1286  #molecular-weight 142915  #checksum 1555

  DB  9;  Score    2806;  Match 40.7%;  QryMatch 28.3%;  Pred. No. 0.00e+00;
  Matches   475;  Conservative  301;  Mismatches 301;  Indels  90;  Gaps  53;

Db     37 aldqidkgkargsrtaiylrsvfqshletlgssvqkhagkvlfvailvlstfcvglksaq 96
Qy     47 ALEQISKGKATGRKAPLWLRAKFQRLLFKLGCYIQKNCGKFLVVGLLIFGAFAVGLKAAN 106

Db     97 ihskvhqlwiqeggrleaelaytqktigedesathqlliqtthdpnasvlhpqallahle 156
Qy    107 LETNVEELWVEVGGRVSRELNYTRQKIGEEAMFNPQLMIQTPKEEGANVLTTEALLQHLD 166

Db    157 vlvkatavkvhlydtewglrdmcnmpstpsfegiyyieqilrhlipcsiitpldcfwegs 216
Qy    167 SALQASRVHVYMYNRQWKLEHLCYKSGELITETGY-MDQIIEYLYPCLIITPLDCFWEGA 225

Db    217 qllgpesavvipglnqrllwttlnpasvmqymkqkmseekisfdfetveqymkraaigsg 276
Qy    226 KLQS-GTAYLL-G-KPPLRWTNFDP---LEFLEE-LK--KINYQVDSWEEMLNKAEVGHG 276

Db    277 ymekpclnplnpncpdtapnknstqppdvgailsggcygyaakhmhwpeelivggrkrnr 336
Qy    277 YMDRPCLNPADPDCPATAPNKNSTKPLDVALVLNGGCQGLSRKYMHWQEELIVGGTVKNA 336

Db    337 sghlrkaqalqsvvqlmtekemydqwqd-nykvhhlgwtqekaaevlnawqrnfsreveq 395
Qy    337 TGKLVSAHALQTMFQLMTPKQMYEHFRGYDY-VSHINWNEDRAAAILEAWQRTYV-EVVH 394

Db    396 llrkqsriatnydiyvfssaalddilakfshpsalsivigvavtvlyafc-tllrwrdpv 454
Qy    395 --QSVAPNSTQ-KVLPFTTTTLDDILKSFSDVSVIRVASGYLLMLAYA-CLTMLRW-DCS 449

Db    455 rgqssvgvagvllmcfstaaglglsallgivfnaastqvvpflalglgvdhifmltaaya 514
Qy    450 KSQGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFS 509

Db    515 es--nrr---eq-tklilkkvgpsilfsacstagsffaaafipvpalkvfclqaaivmcs 568
Qy    510 ETGQNKRIPFEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVVF 569

Db    569 nlaaallvfpamisldlrrrtagradifcc----c----f---p-vwke-qpkv--app- 612
Qy    570 NFAMVLLIFPAILSMDLYRREDRRLDIFCCLTSPCVSRVIQVEPQAYTEPHSNTRYSPPP 629

Db    613 --v-lplnnn-n-g-rga---r---hpks---cnn-n-r----v-plp-aqnpll-e--q 646
Qy    630 PYTSHSFAHETHITMQSTVQLRTEYDPHTHVYYTTAEPRSEISVQPVTVTQDNLSCQSPE 689

Db    647 -----r---adipgss-hsl--a-s-fslatfafqhytpflmrswvkfltvmgflaalis 693
Qy    690 STSSTRDLLSQFSDSSLHCLEPPCTKWTLSSFAEKHYAPFLLKPKAKVVVILLFLGLLGV 749

Db    694 slyastrlqdgldiidlvpkdsnehkfldaqtrlfgfysmyavtqgnfeyptqqqllrdy 753
Qy    750 SLYGTTRVRDGLDLTDIVPRETREYDFIAAQFKYFSFYNMYIVTQKA-DYPNIQHLLYDL 808

Db    754 hdsfvrvphviknd ngglpdfwlllfsewlgnlqkifdeeyrdgrltkecwfpnassdai 813
Qy    809 HKSFSNVKYVMLEENKQLPQMWLHYFRDWLQGLQDAFDSDWETGRIMPNN-YKNGSDDGV 867

Db    814 layklivqtghvdnpvdkelvltnrlvnsdgiinqrafynylsawatndvfaygasqgkl 873
Qy    868 LAYKLLVQTGSRDKPIDISQLTKQRLVDADGIINPSAFYIYLTAWVSNDPVAYAASQANI 927
```



```
Db    874 ypeprqyfhqpney--d--lkipkslplvyaqmpfylhgltdtsqiktlighirdlsvky 929
Qy    928 RPHRPEWVHDKADYMPETRLRIPAAEPIEYAQFPFYLNGLRDTSDFVEAIEKVRVICNNY 987

Db    930 egfglpnypsgipfifweqymtlrsslamilacvllaalvlvsllllsvwaavlvilsvl 989
Qy    988 TSLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAVFLLNPWTAGIIVM-VL 1046

Db    990 asla-qifgamtllgiklsaipavililsvgmmlcfnvlislgfmtsvgnrqrrvqlsmq 1048
Qy   1047 ALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVHVALAFLTAIGDKNHRAMLALE 1106

Db   1049 mslgplvhgmltsgvavfmlstspfefvirhfcwlllvvlcv-gacnsllvfpillsmvg 1107
Qy   1107 HMFAPVLDGAVSTLLGVLMLAGSEFDFIVRYF-FAVLAILTVLGVLNGLVLLPVLLSFFG 1165

Db   1108 peaelvplehpdristpsplpvrsskr 1134
Qy   1166 PCPEVSPANGLNRLPTPSPEPPPSVVR 1192


RESULT    2
ENTRY           S06119     #type complete
TITLE           membrane protein patched - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                  09-Sep-1994
ACCESSIONS      S06119
REFERENCE       S06119
   #authors     Nakano, Y.; Guerrero, I.; Hidalgo, A.; Taylor, A.; Whittle,
                  J.R.S.; Ingham, P.W.
   #journal     Nature (1989) 341:508-513
   #title       A protein with several possible membrane-spanning domains
                  encoded by the Drosophila segment polarity gene patched.
   #cross-references MUID:90015164
   #accession   S06119
      ##status       not compared with conceptual translation
      ##molecule_type DNA
      ##residues     1-1299 ##label NAK
GENETICS
   #gene        ptc
   #map_position 2 44D3-D4
KEYWORDS        glycoprotein; transmembrane protein
FEATURE
   74-92             #domain transmembrane #status predicted #label TM01\
   427-448           #domain transmembrane #status predicted #label TM02\
   456-503           #domain transmembrane #status predicted #label TM03\
   529-555           #domain transmembrane #status predicted #label TM04\
   557-585           #domain transmembrane #status predicted #label TM05\
   677-699           #domain transmembrane #status predicted #label TM06\
   967-1017          #domain transmembrane #status predicted #label TM07\
   1019-1047         #domain transmembrane #status predicted #label TM08\
   1061-1086         #domain transmembrane #status predicted #label TM09\
   1093-1121         #domain transmembrane #status predicted #label TM10\
   142,298,335,388,
   807,861,1194,1271  #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY         #length 1299  #molecular-weight 144091  #checksum 7740

  DB  9;  Score    2703;  Match 40.1%;  QryMatch 27.3%;  Pred. No. 0.00e+00;
  Matches   471;  Conservative  304;  Mismatches 306;  Indels  95;  Gaps  56;
```

```
Db     37 aldqidkgkargsrtaiylrsvfqshletlgssvqkhagkvlfvailvlstfcvglksaq 96
Qy     47 ALEQISKGKATGRKAPLWLRAKFQRLLFKLGCYIQKNCGKFLVVGLLIFGAFAVGLKAAN 106

Db     97 ihskvhqlwiqegggleaelaytqktigedesathqlliqtthdpnasvlhpqallahle 156
Qy    107 LETNVEELWVEVGGRVSRELNYTRQKIGEEAMFNPQLMIQTPKEEGANVLTTEALLQHLD 166

Db    157 vlvkatavkvhlydtewglrdmcnmpstpsfegiyyieqilrhlipcsiitpldcfwegs 216
Qy    167 SALQASRVHVYMYNRQWKLEHLCYKSGELITETGY-MDQIIEYLYPCLIITPLDCFWEGA 225

Db    217 qllgpesavvipglnqrllwttlnpasvmqymkqkmseekisfdfetveqymkraaiasg 276
Qy    226 KLQS-GTAYLL-G-KPPLRWTNFDP---LEFLEE-LK--KINYQVDSWEEMLNKAEVGHG 276

Db    277 ymekpclnplnpncpdtapnknstqppdvgailsggcygyaakhmhwpeelivggakrnr 336
Qy    277 YMDRPCLNPADPDCPATAPNKNSTKPLDVALVLNGGCQGLSRKYMHWQEELIVGGTVKNA 336

Db    337 sghlrkaqalqsvvqlmtekemydqwqd-nykvhhlgwtqekaaevlnawqrnfsreveq 395
Qy    337 TGKLVSAHALQTMFQLMTPKQMYEHFRGYDY-VSHINWNEDRAAAILEAWQRTYV-EVVH 394

Db    396 llrkqsriatnydiyvfssaalddilakfshpsalsivigvavtvlyafc-tllrwrdpv 454
Qy    395 --QSVAPNSTQ-KVLPFTTTTLDDILKSFSDVSVIRVASGYLLMLAYA-CLTMLRW-DCS 449

Db    455 rgqssvgvagvllmcfstaaglglsallgivfnaastqvvpflalglgvdhifmltaaya 514
Qy    450 KSQGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFS 509

Db    515 es--nrr---eq-tklilkkvgpsilfsacstagsffaaafipvpalkvfclqaaivmcs 568
Qy    510 ETGQNKRIPFEDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVVF 569

Db    569 nlaaallvfpamisldlrrrtagradifcc----c----f---p-vwke-qpkv--app- 612
Qy    570 NFAMVLLIFPAILSMDLYRREDRRLDIFCCLTSPCVSRVIQVEPQAYTEPHSNTRYSPPP 629

Db    613 --v-lplnnn-n-g-rga---r---hpks---cnn-n-r--val-p---aqnpll-e--q 646
Qy    630 PYTSHSFAHETHITMQSTVQLRTEYDPHTHVYYTTAEPRSEISVQPVTVTQDNLSCQSPE 689

Db    647 -----r---adipgss-hsl--a-s-fslatfafqhytpflmrswvkfltvmgflaalis 693
Qy    690 STSSTRDLLSQFSDSSLHCLEPPCTKWTLSSFAEKHYAPFLLKPKAKVVVILLFLGLLGV 749

Db    694 slyastrlqdgldiidlvpkdsnehkfldaqtrlfgfysmyavtqgnfeyptqqqllrdy 753
Qy    750 SLYGTTRVRDGLDLTDIVPRETREYDFIAAQFKYFSFYNMYIVTQKA-DYPNIQHLLYDL 808

Db    754 hdsfvrvphviknddngglpdfwlllfsewlgnlqkifdeeyrdgrltkecwfpnassdai 813
Qy    809 HKSFSNVKYVMLEENKQLPQMWLHYFRDWLQGLQDAFDSDWETGRIMPNN-YKNGSDDGV 867

Db    814 laylkivqtghvdnpvdkelvltnrlvnsdgiinqrafynylsawatnassptellranc 873
Qy    868 LAYKLLVQTGSRDKPIDISQLTKQRLVDADGIINPSAFYIYLTAWVSNDPVAYAASQAN- 926

Db    874 irnrangasqgklypeprqyfhqpneydlkipkslplvyaqmpfylhgltdtsqiktlig 933
Qy    927 IRPHRPEWVHDK--AD---YM--P-ETRLRIPAAEPIEYAQFPFYLNGLRDTSDFVEAIE 978
```



```
Db    934 hirdlsvkyegfglpnypsgipfifweqymtlrsslamilacvllaalvlvsllllsvwa 993
Qy    979 KVRVICNNYTSLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAVFLLNPWT 1038

Db    994 avlvilsvlasla-qifgamtllgiklsaipavililsvgmmlcfnvlislgfmtsvgnr 1052
Qy   1039 AGIIVM-VLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVHVALAFLTAIGDK 1097

Db   1053 qrrvqlsmqmslgplvhgmltsgvavfmlstspfefvirhfcwlllvvlcv-gacnslv 1111
Qy   1098 NHRAMLALEHMFAPVLDGAVSTLLGVLMLAGSEFDFIVRYF-FAVLAILTVLGVLNGLVL 1156

Db   1112 fpillsmvgpeaelvplehpdristpsplpvrsskr 1147
Qy   1157 LPVLLSFFGPCPEVSPANGLNRLPTPSPEPPPSVVR 1192


RESULT    3
ENTRY             S52525     #type complete
TITLE             hypothetical protein - yeast (Saccharomyces cerevisiae)
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-May-1995 #sequence_revision 08-May-1995 #text_change
                    08-May-1995
ACCESSIONS        S52525
REFERENCE         S52519
   #authors       Badcock, K.; Churcher, C.
   #submission    submitted to the EMBL Data Library, February 1995
   #accession     S52525
      ##status         preliminary
      ##residues       1-1170 ##label BAD
      ##cross-references EMBL:Z48483
SUMMARY           #length 1170  #molecular-weight 132644  #checksum 5191

  DB 11;  Score     410;  Match 30.2%;  QryMatch 4.1%;  Pred. No. 7.72e-36;
  Matches    62;  Conservative   71;  Mismatches  63;  Indels   9;  Gaps   8;

Db    956 lrsqkdfiqaysdgvr-isssfpeldmfaysp-f-yiffvqyqtlgpltlkligsaiili 1012
Qy    967 LRDTSDFVEAI-EKVRVICNNYTSLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACT 1025

Db   1013 ffissvflqnirssfllalvvtmiivdigalmallgislnavslvnliicvglgvefcvh 1072
Qy   1026 FLVCAVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVH 1085

Db   1073 ivrsftvvpsetkkdansrvlyslntigesvikgitltkfigvcvlafaqskifdvfyfr 1132
Qy   1086 VALAFLTAIGD--KN-H-RAMLALEHMFAPVLDG-AVSTLLGVLMLAGSEFDFIVRYFFA 1140

Db   1133 mwftliivaalhallflpallslfg 1157
Qy   1141 VLAILTVLGVLNGLVLLPVLLSFFG 1165


RESULT    4
ENTRY             S44795     #type complete
TITLE             F09G8.3 protein - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              14-Sep-1994 #sequence_revision 12-May-1995 #text_change
                    12-May-1995
ACCESSIONS        S44795
REFERENCE         S44774
   #authors       Anderson, K.
   #submission    submitted to the EMBL Data Library, February 1993
   #description   Sequence of the C. elegans cosmid F09G8.
```

```
    #accession    S44795
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-633 ##label AND
      ##cross-references EMBL:L11247
  GENETICS
    #introns       38/1; 111/1; 142/3; 180/3; 241/3; 262/2; 294/2; 322/2; 409/2;
                     495/2; 531/2; 569/2; 604/1
  SUMMARY          #length 633  #molecular-weight 70659  #checksum 1119

   DB  9;  Score     236;  Match 27.8%;  QryMatch 2.4%;  Pred. No. 4.06e-13;
   Matches    50;  Conservative   49;  Mismatches  76;  Indels   5;  Gaps   4;

Db     58 gvkvyvystffpyyeqyltlsttvytlvvlvlfvafvtislflrvn-lagslvtvfvlls 116
          |:  |  :  | : |||::|   :   : :||  :|:  ::||     || :| |: |::
Qy    991 GLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAVFLLNPWTAGIIVMVLALMT 1050

Db    117 sylhlmewmyllgitvnvvsvinmamslgiaveffgqmlhgfynskkpkreerafaalvs 176
            |  |  | |:|| :: |:|: :  |:|||:|||  ::  :|  :   | : ||  ||
Qy   1051 VELFGM--MGLIGIKLSAVPVVILIASVGIGVEFTVHVALAFLTAIGDK-NHRAMLALEH 1107

Db    177 ngsttlsgifpaimitagclsfadsrvlityfcnqlvgiglvcavhgvvymptllaifgs 236
            :  | |  ::::     |: ::   :: ||   |  : ::  ::|:| :| ||::||:
Qy   1108 MFAPVLDGAVSTLLGVLM-LAGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLSFFGP 1166


  RESULT    5
  ENTRY            S22697     #type fragment
  TITLE            extensin - Volvox carteri (fragment)
  ORGANISM         #formal_name Volvox carteri
  DATE             04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
                     18-Jun-1993
  ACCESSIONS       S22697; S21006
  REFERENCE        S22697
    #authors       Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
    #journal       EMBO J. (1992) 11:2055-2062
    #title         A novel extensin that may organize extracellular matrix
                     biogenesis in Volvox carteri.
    #cross-references MUID:92289669
    #accession    S22697
      ##molecule_type mRNA
      ##residues      1-464 ##label HAL
      ##cross-references EMBL:X65165
  KEYWORDS         glycoprotein
  SUMMARY          #length 464  #checksum 3143

   DB  8;  Score     145;  Match 25.1%;  QryMatch 1.5%;  Pred. No. 8.31e-03;
   Matches    44;  Conservative   42;  Mismatches  82;  Indels   7;  Gaps   6;

Db    219 tpsppppprvstspppparvssspppatrs-ppprritspspvltaspplpktsppppr 277
          |||| ||| |   : ||:: :::: ::    ::   ::: |  |    :   :: |:
Qy   1181 TPSPEPPPSVVRFAVPPGHTNNGSDSSDSEYSSQTTVSGISEELRQYEAQQGAGGPAHQV 1240

Db    278 vppsppppv-a-sppppppprvsps-ppppqpvsspppppppprpspspp--pprsspspp 332
          :  :   || | |    | :|  |:  |  ||   : : :| | :  |   ||| :  ||
Qy   1241 IVEATENPVFARSTVVHPDSRHQPPLTPRQQPHLDSGSLSPGRQGQQPRRDPPREGLRPP 1300

Db    333 ppsppppsppppprpspsppppprsspspppp vvsp-pppppraspppppasspppp 386
          |  |   :   :  : | |: |  ::|  ||::      | : ::  ||:::|:|
Qy   1301 PYRPRRDAFEISTEGHSGPSNRDRSGPVGPVLTTLGTQRPPPWAALCPATASPSP 1355


  RESULT    6
  ENTRY            S45552     #type complete
```



```
  TITLE            penicillin-binding protein 5* precursor (version 2) -
                     Bacillus subtilis
  ORGANISM         #formal_name Bacillus subtilis
  DATE             06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                     24-Feb-1995
  ACCESSIONS       S45552
  REFERENCE        S45533
    #authors       Sorokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.;
                     Serror, P.
    #submission    submitted to the EMBL Data Library, November 1993
    #accession     S45552
      ##molecule_type DNA
      ##residues      1-382 ##label SOR
      ##cross-references EMBL:L09228
  COMMENT     From position 179 to 194, this sequence differs by a frame shift
                from the sequence in entry PIR2:B42274.
  SUMMARY          #length 382  #molecular-weight 43081  #checksum 5757

   DB  7;  Score     143;  Match 28.0%;  QryMatch 1.4%;  Pred. No. 1.32e-02;
   Matches    21;  Conservative   22;  Mismatches  28;  Indels   4;  Gaps   3;

Db    233 staskdgidliavtindpndwddhmkmfnyvfehyqtyliakkgdipklkgtfye-skaf 291
          :|  :||:|| :   :  : |     |:| |  |  |::: |:| |::   :|:  |:|
Qy    754 TTRVRDGLDLTDIVPRETREYDFIAAQFKY-FSFYNMYIVTQKADYPNIQHLLYDLHKSF 812

Db    292 ikrdityllteeeke 306
            :: |:: ||:|:
Qy    813 --SNVKYVMLEENKQ 825


  RESULT    7
  ENTRY            B42274     #type complete
  TITLE            penicillin-binding protein 5* precursor (version 1) -
                     Bacillus subtilis
  ALTERNATE_NAMES  PBP 5*
  CONTAINS         serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4)
  ORGANISM         #formal_name Bacillus subtilis
  DATE             04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                     24-Feb-1995
  ACCESSIONS       B42274
  REFERENCE        A42274
    #authors       Buchanan, C.E.; Ling, M.L.
    #journal       J. Bacteriol. (1992) 174:1717-1725
    #title         Isolation and sequence analysis of dacB, which encodes a
                     sporulation-specific penicillin-binding protein in Bacillus
                     subtilis.
    #cross-references MUID:92193254
    #accession     B42274
      ##status        not compared with conceptual translation
      ##molecule_type DNA; protein
      ##residues      1-382 ##label BUC
      ##cross-references NCBIP:87676
      ##note          sequence extracted from NCBI backbone
  COMMENT     This penicillin-binding protein is expressed specifically in the
                mother cell compartment, starting from stage III of sporulation.
  COMMENT     From position 179 to 194, this sequence differs by a frame shift
                from the sequence in entry PIR2:S45552.
  GENETICS
    #gene          dacB
  KEYWORDS         hydrolase; membrane protein; serine carboxypeptidase;
                     sporulation
  FEATURE
    1-27               #domain signal sequence #status predicted #label SIG
  SUMMARY          #length 382  #molecular-weight 43120  #checksum 5082
```

```
  DB  7;  Score     143;  Match 28.0%;  QryMatch 1.4%;  Pred. No. 1.32e-02;
  Matches    21;  Conservative   22;  Mismatches  28;  Indels   4;  Gaps   3;

Db    233 staskdgidliavtindpndwddhmkmfnyvfehyqtyliakkgdipklkgtfye-skaf 291
          :|  :||:||  :   :  : |     |:| |  |  |::: |:| |::   :|:  |:|
Qy    754 TTRVRDGLDLTDIVPRETREYDFIAAQFKY-FSFYNMYIVTQKADYPNIQHLLYDLHKSF 812

Db    292 ikrdityllteeeke 306
            :: |:: ||:|:
Qy    813 --SNVKYVMLEENKQ 825


RESULT    8
ENTRY             C40046     #type complete
TITLE             antibiotic transport-associated protein actII-3 -
                    Streptomyces coelicolor
ORGANISM          #formal_name Streptomyces coelicolor
DATE              30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
                    18-Jun-1993
ACCESSIONS        C40046
REFERENCE         A40046
   #authors       Fernandez-Moreno, M.A.; Caballero, J.L.; Hopwood, D.A.;
                    Malpartida, F.
   #journal       Cell (1991) 66:769-780
   #title         The act cluster contains regulatory and antibiotic export
                    genes, direct targets for translational control by the bldA
                    tRNA gene of Streptomyces.
   #cross-references MUID:91347376
   #accession     C40046
      ##molecule_type DNA
      ##residues      1-711 ##label FER
      ##cross-references GB:M64683
SUMMARY           #length 711  #molecular-weight 74862  #checksum 4136

  DB  8;  Score     137;  Match 20.1%;  QryMatch 1.4%;  Pred. No. 5.23e-02;
  Matches    31;  Conservative   57;  Mismatches  60;  Indels   6;  Gaps   5;

Db    184 llivvtilvvtyrspllwllpmisagm-slvisqaivyllaknagltvn-aqtamiltvl 241
          || :  :|  |:    : ||   :||:  :|::   | |::: : : :: :  :::: :
Qy   1015 LLSISVVLACTFLVCAVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIKLSAVPVVILIA 1074

Db    242 vlgaatdyalllvaryreelrrhedrheamavalrragpaivasaatvavsmlvlllaal 301
           :| : :::: :   :   :   :  | || :|| :  :::: :| :  :::|:|  : :
Qy   1075 SVGIGVEFTVHVALAFLTAIG--DKNHRAM-LALEHMFAPVLDGAVSTLLGVLMLAGSEF 1131

Db    302 n-stkglgpvcavgvlvgllsmmtllpallvifg 334
          :   : : :| |:  ::|:|: : ||| || :||
Qy   1132 DFIVRYFFAVLAILTVLGVLNGLVLLPVLLSFFG 1165


RESULT    9
ENTRY             BVBRCB     #type complete
TITLE             cyaB protein - Bordetella pertussis
ORGANISM          #formal_name Bordetella pertussis
DATE              31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
                    08-Dec-1994
ACCESSIONS        S02386
REFERENCE         S02386
   #authors       Glaser, P.; Sakamoto, H.; Bellalou, J.; Ullmann, A.; Danchin,
                    A.
   #journal       EMBO J. (1988) 7:3997-4004
   #title         Secretion of cyclolysin, the calmodulin-sensitive adenylate
                    cyclase--haemolysin bifunctional protein of Bordetella
```



```
                    pertussis.
   #cross-references MUID:89091151
   #accession     S02386
      ##molecule_type DNA
      ##residues      1-712 ##label GLA
COMMENT           This protein is required for the transport of cyclolysin (or
                    calmodulin-sensitive adenylate cyclase--hemolysin bifunctional
                    protein, encoded by cyaA) across the cell envelope and for its
                    release into the external medium. This secretion process is very
                    similar to that of the E. coli alpha-hemolysin.
GENETICS
   #gene          cyaB
CLASSIFICATION    #superfamily hemolysin secretion protein B; malK protein
                    homology
KEYWORDS          ATP binding; cyclolysin transport; membrane protein; P-loop
FEATURE
   488-682             #domain malK protein homology #label MK1\
   505-513             #region nucleotide-binding motif A (P-loop)\
   629-633             #region nucleotide-binding motif B\
   511                 #binding_site ATP (Lys) #status predicted
SUMMARY           #length 712  #molecular-weight 77969  #checksum 2892

  DB  3;  Score     137;  Match 27.6%;  QryMatch 1.4%;  Pred. No. 5.23e-02;
  Matches    43;  Conservative   39;  Mismatches  62;  Indels  12;  Gaps  10;

Db    166 lvlqfislltplffqvvmdkvlvnnametlnvitvgflaailfeal-ltgirtylfahts 224
          |:| :  :|:  |: |     |:|     : |: :::::  ||  : | ||:  | |
Qy   1014 LLLSISVVLACTFL-VC-AVFLLNPWTAGIIVMVLALMTVELFGMMGLIGIK--LSAVPV 1069

Db    225 skldvelgarlyahllrlplayfqarrvgdsvarvr-elehirafltgnavtvlldv-vf 282
            |   :|  :    ::::||:: |  :||   |    |||: | :   ||: || | ::
Qy   1070 VILIASVGIGV-EFTVHVALAFLTA--IGDKNHRAMLALEHMFAPVLDGAVSTLLGVLML 1126

Db    283 svv-f-iavmffysvkltlvvlaalpcyfllslvlt 316
          :   | : | :|::|   | ||: |    ||:::|:
Qy   1127 AGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLS 1162


RESULT   10
ENTRY             S16564     #type complete
TITLE             nolH protein - Rhizobium meliloti
ORGANISM          #formal_name Rhizobium meliloti
DATE              13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                    13-Jan-1995
ACCESSIONS        S16564
REFERENCE         S16561
   #authors       Baev, N.; Endre, G.; Petrovics, G.; Banfalvi, Z.; Kondorosi,
                    A.
   #journal       Mol. Gen. Genet. (1991) 228:113-124
   #title         Six nodulation genes of nod box locus 4 in Rhizobium meliloti
                    are involved in nodulation signal production: nodM codes
                    for D-glucosamine synthetase.
   #cross-references MUID:91360053
   #accession     S16564
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-215 ##label BAE
      ##cross-references EMBL:X58632
SUMMARY           #length 215  #molecular-weight 23775  #checksum 2405

  DB  7;  Score     135;  Match 23.4%;  QryMatch 1.4%;  Pred. No. 8.20e-02;
  Matches    32;  Conservative   47;  Mismatches  49;  Indels   9;  Gaps   8;

Db      1 mfl-nswrstvitgltlpisvigtfaaiyalgftlnimtlmalslsigiliddtivvren 59
```

```
          :|| |:| : :|  : |::  :  |: :  :|: |: : :: |  |:|| :: |: |
Qy   1031 VFLLNPWTAGIIV-MVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFTVHVALA 1089

Db     60 itrhlqmgkdpvraaldgtneiglavlsttlc-ivavflpvafmggligrfflqfgvtva 118
          :   :   |:  || |   : :: :||  ::  :::|::  :    :| |:|  |:| :|
Qy   1090 FLTAIG-DKNH-RAMLALEHMFA-PVLDGAVSTLLGVLMLAGSEFDFIVRYF--FAV-LA 1143

Db    119 vavvislfvsftldpml 135
          :  |:::: :: | |:|
Qy   1144 ILTVLGVLNGLVLLPVL 1160


RESULT   11
ENTRY            S28276     #type complete
TITLE            hypothetical protein F54G8.5 - Caenorhabditis elegans
ORGANISM         #formal_name Caenorhabditis elegans
DATE             12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
                   30-Sep-1993
ACCESSIONS       S28276
REFERENCE        S28273
   #authors      Sulston, J.
   #submission   submitted to the EMBL Data Library, December 1992
   #accession    S28276
      ##molecule_type DNA
      ##residues      1-413 ##label SUL
      ##cross-references EMBL:Z19155
GENETICS
   #introns      46/1; 95/1; 210/2; 339/3
SUMMARY          #length 413  #molecular-weight 46966  #checksum 9391

  DB  9;  Score     134;  Match 29.6%;  QryMatch 1.4%;  Pred. No. 1.03e-01;
  Matches    24;  Conservative   24;  Mismatches  31;  Indels   2;  Gaps   2;

Db      9 eqryihaltesaaslfltsltdglsfaigsisdfhavrvfctycamailfmflfqvtffn 68
          |:|    |  ::||: |||:::  :| ::::  : |:| |    |: ::| |   : :|
Qy    520 EDRTGECLKRTGASVALTSISNVTAFFMAALIPIPALRAFSLQAAVVVVFNFAMVLLIFP 579

Db     69 avmslc-crrevsgkhpvfcc 88
          |::|:   |||   :  :|||
Qy    580 AILSMDLYRREDR-RLDIFCC 599


RESULT   12
ENTRY            B26430     #type complete
TITLE            arabinose transport protein - Escherichia coli
ORGANISM         #formal_name Escherichia coli
DATE             05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change
                   27-Jan-1995
ACCESSIONS       B26430; A28075
REFERENCE        A93389
   #authors      Maiden, M.C.J.; Davis, E.O.; Baldwin, S.A.; Moore, D.C.M.;
                   Henderson, P.J.F.
   #journal      Nature (1987) 325:641-643
   #title        Mammalian and bacterial sugar transport proteins are
                   homologous.
   #cross-references MUID:87115869
   #accession    B26430
      ##status        not compared with conceptual translation
      ##molecule_type DNA
      ##residues      1-472 ##label MAI
      ##note          nucleotide sequence is not given
REFERENCE        A28075
   #authors      Maiden, M.C.J.; Jones-Mortimer, M.C.; Henderson, P.J.F.
   #journal      J. Biol. Chem. (1988) 263:8003-8010
```



```
   #title        The cloning, DNA sequence, and overexpression of the gene
                   araE coding for arabinose-proton symport in Escherichia
                   coli K12.
   #cross-references MUID:88228015
   #accession    A28075
      ##molecule_type DNA
      ##residues      1-472 ##label MA2
GENETICS
   #gene         araE
   #map_position 61 min
KEYWORDS         arabinose transport; membrane protein; symport system
SUMMARY          #length 472  #molecular-weight 51684  #checksum 3179

  DB  7;  Score     130;  Match 23.6%;  QryMatch 1.3%;  Pred. No. 2.49e-01;
  Matches    42;  Conservative   61;  Mismatches  56;  Indels  19;  Gaps  17;

Db     37 gldigviagalpfit-dhfv-ltsrlq-e-wvv-ssmmlgaaigalfngwlsfrlgrkys 91
          :| ::  : : ||:  :::: |  :|     || :   |  |:  |:| | :  :    :
Qy    989 SLGLSSYPNGYPFLFWEQYISLRHWLLLSISVVLACTFLVCAV-FLLNPWTAGIIVMVLA 1047

Db     92 lmagailfvlgsigsafat-svemliaarvvlgiavgiasytaplylsemasenvrgkm- 149
          ||:  :: ::| ||  ::: :| :|||:  | ||:| ::   |  :|: ::  | |:::
Qy   1048 LMTVELFGMMGLIGIKLSAVPVVILIAS--V-GIGVEFTVHVALAFLTAIGDKNHRAMLA 1104

Db    150 is-my-qlm-vtlgivla-f-lsdtafsysgnwr-amlgvlalpavl--liilvvflp 199
          :  |:  ::  ::: :|: : |: : | :      |:|::|:: :||  |::| |:|:
Qy   1105 LEHMFAPVLDGAVSTLLGVLMLAGSEFDFIVRYFFAVLAILTVLGVLNGLVLLPVLLS 1162


RESULT   13
ENTRY            S52796     #type complete
TITLE            prpL_2 protein - human (fragment)
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             19-May-1995 #sequence_revision 19-May-1995 #text_change
                   19-May-1995
ACCESSIONS       S52796
REFERENCE        S52796
   #authors      Ruhlmann, A.; Kreideweiss, S.; Nordheim, A.
   #submission   submitted to the EMBL Data Library, March 1995
   #accession    S52796
      ##status        preliminary
      ##residues      1-403 ##label RUH
      ##cross-references EMBL:X86019
SUMMARY          #length 403  #molecular-weight 42083  #checksum 6830

  DB 12;  Score     130;  Match 26.6%;  QryMatch 1.3%;  Pred. No. 2.49e-01;
  Matches    50;  Conservative   40;  Mismatches  90;  Indels   8;  Gaps   8;

Db     51 sppsgpgrfpvpspghrsgppepqrnrmppprpdvgskpdsipppvpstprpiqsslhnr 110
          ||::|  |:| |||   : ||   |  :|| : : || :     :  :|   |   |:
Qy   1171 SPANGLNRLPTPSP-E-P-PPSVVRFAVPPGHTNNGSDSSDSEYSSQTTVSGISEELRQY 1227

Db    111 gsppvpggp-rqpspgptpppfpgnrgtalgggsirqsplsssspfsnrpplpptpsral 169
           :  :||| :|   :|  |  : |:| :   | :|:||::   ::  : : :|:|
Qy   1228 EAQQGAGGPAHQVIVEATENPVFA-RSTVVHPDSRHQPPLTPRQQ-PHLDSGSLSPGRQ- 1284

Db    170 ddkpppppppvgnrpsihreavpppppqnnkppvpstprpsaphrphlrpppprpgp-p 228
           : |  ||  | ||:  |   :     :  : ||   |:|  | |      || | :
Qy   1285 GQQPRRDPPREGLRPPPYRPRRDAFEISTEGHSGPSNRDRSGPVGPVLTTLGTQRPPPWA 1344

Db    229 plppsssg 236
          :| |::::
Qy   1345 ALCPATAS 1352
```

```
RESULT   14
ENTRY            A33830     #type complete
TITLE            cation efflux system membrane protein czcA - Alcaligenes
                   eutrophus
ORGANISM         #formal_name Alcaligenes eutrophus
DATE             23-Mar-1990 #sequence_revision 18-Sep-1992 #text_change
                   31-Dec-1993
ACCESSIONS       A33830
REFERENCE        A33830
   #authors      Nies, D.H.; Nies, A.; Chu, L.; Silver, S.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1989) 86:7351-7355
   #title        Expression and nucleotide sequence of a plasmid-determined
                   divalent cation efflux system from Alcaligenes eutrophus.
   #cross-references MUID:90017477
   #accession    A33830
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-1063 ##label NIE
      ##cross-references GB:M26073
KEYWORDS         membrane protein
SUMMARY          #length 1063  #molecular-weight 115643  #checksum 209

  DB  7;  Score     130;  Match 25.6%;  QryMatch 1.3%;  Pred. No. 2.49e-01;
  Matches    23;  Conservative   30;  Mismatches  34;  Indels   3;  Gaps   3;

Db    457 liimivylpifaltgvegkmfhpmaftvvlallgamilsvtfvpaaval-figervaeke 515
          :|:|:: |    | |: | :   ::   |: |:::: ::| |    ||| |:   :::|:
Qy   1041 IIVMVLALMTVELFGMMGLIGIKLSAVPVVILIASVGIGVEFT-VHVALAFLTA-IGDKN 1098

Db    516 nrlmlwakrryeplleklslantavvltfaa 545
          :| ||   : : |:|: :::    || :|:
Qy   1099 HRAMLALEHMFAPVLDGAVSTLLGVLMLAG 1128


RESULT   15
ENTRY            B48013     #type complete
TITLE            proline-rich proteoglycan 2 precursor, parotid - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change
                   02-Jun-1995
ACCESSIONS       B48013
REFERENCE        A48013
   #authors      Castle, A.M.; Castle, J.D.
   #journal      J. Biol. Chem. (1993) 268:20490-20496
   #title        Novel secretory proline-rich proteoglycans from rat parotid.
                   Cloning and characterization by expression in AtT-20 cells.
   #accession    B48013
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-295 ##label CAS
      ##cross-references GB:L17318
KEYWORDS         extracellular protein; glycoprotein; tandem repeat
SUMMARY          #length 295  #molecular-weight 30026  #checksum 4849

  DB 10;  Score     129;  Match 26.9%;  QryMatch 1.3%;  Pred. No. 3.09e-01;
  Matches    47;  Conservative   34;  Mismatches  85;  Indels   9;  Gaps   9;

Db    101 ppaagpqrppqpgspqgppppggpqqrppqgpppqggpqrppqpgspqgppppggpqqrp 160
          :|| |  | | | ||: |||:      || |    |:    :: :| |   :  : : |
Qy   1171 SPANGLNRLPTP-SPE-PPPSVVRFAVPP-GHTNNGSDSSDSEYSS-QTTVSGISEELRQ 1226

Db    161 pqgpppqggpqrppqpgspqgpp-ppggpqqrapqgpppqggpqrppqpgspqgppppgg 219
          ::    ||| :     : : |  : :   :  :: ||    |:|   ::: :|   |
```



```
Qy   1227 YEAQQGAGGPAHQVIVEATENPVFARSTVVHPDSRHQPPLTPRQQPHLDSGSLSPGRQGQ 1286

Db    220 pqqr-pp-qg-pppqggpqrppqpgspqgppppggpqqrppqgpppqg-gpqrpp 270
            :| || :|  ||   |:| :   | :| : |:  ::  | ||     | ||||
Qy   1287 QPRRDPPREGLRPPPYRPRRDAFEISTEGHSGPSNRDRSGPVGPVLTTLGTQRPP 1341

Search completed: Wed Jan 17 17:22:16 1996
Job time : 191 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:15:17 1996; MasPar time 16.67 Seconds
499.639 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-4
Description: (1:1311) from US08319745.pep
Perfect Score: 9491
Sequence: 1 MVAPDSEAPSNPRITAAHES.....YRDERDHRASPREKRQRFWT 1311

Scoring table: PAM 150
Gap 11

Searched: 53402 seqs, 6354270 residues

Database: a-geneseq18
1 part1
2 part2
3 part3
4 part4
5 part5
6 part6
7 part7
8 part8
9 part9
10 part10

Statistics: Mean 40.660; Variance 210.188; scale 0.193

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	267	2.8	162	7	R37991	Sequence of a new cyt	3.41e-10
2	159	1.7	3910	7	R38470	ALL-1 protein.	1.27e-02
3	159	1.7	1400	8	R44514	MLL amino acid sequen	1.27e-02
4	150	1.6	3969	9	R52971	Product of the cDNA e	4.94e-02
5	144	1.5	765	1	P92275	Human topoisomerase I	1.21e-01
6	144	1.5	265	3	R12844	HTLV-1 protein expres	1.21e-01
7	137	1.4	740	5	R27530	Plasmodium falciparum	3.37e-01
8	133	1.4	123	9	R51053	Epstein-Barr nuclear	6.02e-01

9	133	1.4	1023	3	R13319	Partial Human Natural	6.02e-01
10	129	1.4	183	3	P60624	Sequence B encoded by	1.07e+00
11	129	1.4	751	1	P94776	Novel amyloid precurs	1.07e+00
12	129	1.4	179	6	R32082	Plasmid pSGHEC-1414 p	1.07e+00
13	129	1.4	179	3	R14216	HCV detecting peptide	1.07e+00
14	127	1.3	657	6	R29580	FMR-1 gene product.	1.42e+00
15	124	1.3	1141	6	R31961	Human cardiac cGI PDE	2.17e+00
16	123	1.3	566	3	R20181	Sequence encoded by 6	2.50e+00
17	122	1.3	3025	1	P93284	Sequence of clone HIV	2.88e+00
18	120	1.3	301	2	P70867	Sequence of acidic ba	3.80e+00
19	116	1.2	911	3	R15355	Human erythrocyte mem	6.61e+00
20	114	1.2	1284	1	P81187	Sequence encoded by a	8.70e+00
21	113	1.2	517	4	R22904	1-Caldesmon.	9.97e+00
22	111	1.2	720	3	R15381	Pseudomonas SY77-glut	1.31e+01
23	111	1.2	2237	6	R33550	Sequence of the alpha	1.31e+01
24	111	1.2	720	3	R14445	Pseudomonas SY77-glut	1.31e+01
25	111	1.2	720	3	R15382	Pseudomonas SY77-glut	1.31e+01
26	111	1.2	720	3	R15380	Pseudomonas SY77-glut	1.31e+01
27	111	1.2	561	3	P61363	Soybean glycinin A5A4	1.31e+01
28	111	1.2	720	3	R15379	Pseudomonas SY77-glut	1.31e+01
29	110	1.2	1784	1	R05898	Gene product of first	1.50e+01
30	110	1.2	445	3	R14163	Cellular DNA-binding	1.50e+01
31	109	1.1	493	5	R26944	P.falciparum LSA gene	1.71e+01
32	109	1.1	292	3	P60645	Mouse kidney cell Ban	1.71e+01
33	109	1.1	316	5	R26941	P.falciparum LSA-R-NR	1.71e+01
34	108	1.1	3080	1	P93285	Sequence of clone HIV	1.96e+01
35	107	1.1	462	1	R05766	Portion of peptide an	2.24e+01
36	106	1.1	382	7	R39224	Nucleocapsid protein	2.56e+01
37	106	1.1	318	5	R26943	P.falciparum LSA N-te	2.56e+01
38	105	1.1	3210	1	P81770	Deduced sequence enco	2.92e+01
39	105	1.1	312	5	R27361	Sequence of a polypep	2.92e+01
40	105	1.1	532	5	R27362	Sequence of a polypep	2.92e+01
41	105	1.1	558	5	R27363	Sequence of a polypep	2.92e+01
42	104	1.1	482	3	R21409	NADH dehydrogenase su	3.33e+01
43	103	1.1	1026	9	R48993	rsaA S-lyaer protein.	3.80e+01
44	103	1.1	307	3	R12785	Caldesmon-like polype	3.80e+01
45	103	1.1	1931	5	R27649	Human calcium channel	3.80e+01

ALIGNMENTS

RESULT 1
ID R37991 standard; Protein; 162 AA.
AC R37991;
DT 29-SEP-1993 (first entry)
DE Sequence of a new cytokine which inhibites induction by gamma
DE interferon of expression of Class II histocompatibility antigens.
KW Cytokine; interferon-gamma antagonist; autoimmune disease therapy;
KW transplants; Class II histocompatibility antigens;
KW gamma interferon; leukaemia line K562.
OS Homo sapiens.
PN WO9311232-A.
PD 10-JUN-1993.
PF 02-DEC-1992; F01123.
PR 02-DEC-1991; FR-014908.
PA (INRM) INSERM INST NAT SANTE & RECH MED.
PI Augery-bourget Y, Azzarone B, Boucheix C, Jasmin C;
PI Krief PH;
DR WPI; 93-197051/24.
DR N-PSDB; Q43704.
PT New cytokine as interferon-gamma antagonist - inhibits induction
PT of class II histocompatability antigens on cell surface by
PT IFN-gamma, for treating and preventing auto:immune disorders
PS Claim 1; Fig 1; 44pp; French.

```
CC   The new cytokine specifically inhibits induction by gamma interferon
CC   of expression of Class II histocompatibility antigens on cell
CC   surfaces. It is naturally secreted by cells which are difficult or
CC   impossible to induce with gamme-interferon, e.g. the leukaemia line
CC   K562, from which it can be isolated. To isolate DNA, a K562 gene
CC   bank was constructed in lambda gt11 phage, used to transform Y1090
CC   bacteria and these screened for reactivity with MAbs specific for
CC   the cytokine.Positive colonies were purified,their inserts isolated
CC   with EcoRI and subcloned into pBLSCR KS. Recombinant plasmids were
CC   used to transform HB101 cells and the insert from one positive clone
CC   sequenced. This insert was recloned in vector CDM8 and recombinants
CC   used to transform COS cells. The transformants expressed functional
CC   recombinant cytokine.
SQ   Sequence   162 AA;

  DB  7;  Score     267;  Match 60.8%;  QryMatch 2.8%;  Pred. No. 3.41e-10;
  Matches    31;  Conservative   13;  Mismatches   5;  Indels   2;  Gaps   2;

Db     19 rdkereryrererdrerdrdrdrererererdrererrdr-er-eeekkrhs 67
          || :|:| |||:|||:|||||||:|:|:||||| ||||| :|  :|: :::
Qy   1250 RDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRA 1300


RESULT    2
ID   R38470 standard; Protein; 3910 AA.
AC   R38470;
DT   08-NOV-1993  (first entry)
DE   ALL-1 protein.
KW   Acute lymphoblastic leukemia gene; ALL-1; chromosome 11; treatment;
KW   translocation breakpoint mapping; chromosomal abnormality; diagnosis;
KW   human; acute lymphocytic; myelomonocytic; monocytic; myelogenous;
KW   leukemia; Drosophila; trithorax; homology region; zinc finger domain;
KW   cysteine-rich.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Region          1021..1221
FT   /note= "Region of homology to Drosophila trithorax"
FT   Region          1462..1570
FT   /note= "Region of homology to Drosophila trithorax"
FT   Region          3348..3562
FT   /note= "Region of homology to Drosophila trithorax"
PN   WO9312136-A.
PD   24-JUN-1993.
PF   09-DEC-1992; U10930.
PR   11-DEC-1991; US-805093.
PR   27-MAY-1992; US-888839.
PR   30-OCT-1992; US-971094.
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   Canaani E,  Croce CM;
DR   WPI; 93-214090/26.
DR   N-PSDB; Q43526.
PT   Detection and treatment of acute leukaemia(s) - using prods.
PT   derived from oligo:nucleotide sequences within the ALL-1 gene of
PT   chromosome 11
PS   Disclosure; Page 29-50; 90pp; English.
CC   This sequence is encoded by the acute lymphoblastic leukemia (ALL-1)
CC   gene of chromosome 11.  The ALL-1 gene was isolated by translocation
CC   breakpoint mapping.  Fragments of the ALL-1 cDNA may be used to
CC   identify chromosomal abnormalities within the ALL-1 gene.  These
CC   fragments may be used in the treatment and diagnosis of human
CC   leukemias such as acute lymphocytic, myelomonocytic, monocytic and
CC   myelogenous leukemia.  ALL-1 protein shows three regions of homology
CC   to the Drosophila trithorax protein.  These regions show 64%, 66% and
CC   82% similarity respectively, to the Drosophila gene.  The third region
```



```
CC   of homology constitutes the extreme C-terminus of the two proteins,
CC   both proteins end in an identical sequence.  The first homology region
CC   is cysteine-rich and contains sequence motifs analogous to four zinc
CC   finger domains (3-6) within the trithorax gene.  The second region of
CC   homology is also cysteine-rich and corresponds to zinc fingers 7 and 8
CC   of the Drosophila gene.  The multiple conserved cysteines and
CC   histidines at the 3' end of the motifs allow two or three arrangements
CC   of the putative fingers.  The structure of these cysteine-rich domains
CC   appears to be unique to the trithorax and ALL-1 genes.
SQ   Sequence   3910 AA;

  DB  7;  Score     159;  Match 28.8%;  QryMatch 1.7%;  Pred. No. 1.27e-02;
  Matches    17;  Conservative   22;  Mismatches   18;  Indels   2;  Gaps   2;

Db    784 ergrnkdkapeelskdrdadksvekdksrerdrerekenkre-srkekrkkgseiqsss 841
          :| |::|   :   ::|| |:  ::|: |:|||:|::|  ||  |::: :   : ::|:
Qy   1245 DRRRDRDEDRDR-DRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASP 1302


RESULT    3
ID   R44514 standard; Protein; 1400 AA.
AC   R44514;
DT   28-JUN-1994  (first entry)
DE   MLL amino acid sequence.
KW   Leukaemia; leukemia; MLL gene rearrangement; detection; centromeric;
KW   11q23 chromosome; translocation; breakpoint region.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Region          323..623
FT   /note= "centromeric to the breakpoint region"
FT   Region          574..810
FT   /note= "zinc finger region"
FT   Region          1057..1184
FT   /note= "zinc finger region"
PN   WO9325713-A.
PD   23-DEC-1993.
PF   17-JUN-1993; U05857.
PR   17-JUN-1992; US-900689.
PR   16-DEC-1992; US-991244.
PA   (ARCH-) ARCH DEV CORP.
PI   Diaz MO, Rowley JD.
DR   WPI; 94-007568/01.
PT   Detecting MLL gene rearrangements and translocation - by using
PT   nucleic acid probes, for diagnosing leukaemia
PS   Claim 28; Page 103-111; 136pp; English.
CC   The sequence is that of an MLL amino acid sequence centromeric
CC   to the MLL translocation breakpoint region.
SQ   Sequence   1400 AA;

  DB  8;  Score     159;  Match 28.8%;  QryMatch 1.7%;  Pred. No. 1.27e-02;
  Matches    17;  Conservative   22;  Mismatches   18;  Indels   2;  Gaps   2;

Db     68 ergrnkdkapeelskdrdadksvekdksrerdrerekenkre-srkekrkkgseiqsss 125
          :| |::|   :   ::|| |:  ::|: |:|||:|::|  ||  |::: :   : ::|:
Qy   1245 DRRRDRDEDRDR-DRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASP 1302


RESULT    4
ID   R52971 standard; Protein; 3969 AA.
AC   R52971;
DT   27-SEP-1994  (first entry)
DE   Product of the cDNA encoding htrx.
KW   Human; trithorax gene; L01986; diagnosis; treatment;
KW   immunodeficiency; developmental abnormalities; inherited diseases;
```

KW cancer; acute lymphocytic leukaemia; myelomonocytic leukaemia.
OS Homo sapiens.
PN WO9407502-A.
PD 14-APR-1994.
PF 24-SEP-1993; U09087.
PR 30-SEP-1992; US-954112.
PR 13-MAY-1993; US-061376.
PA (SALK) SALK INST BIOLOGICAL STUDIES.
PI Djabaki M, Evans GA, Parry P, Selleri L;
DR WPI; 94-135206/16.
PT Nucleic acid encoding a human tri:thorax protein - used to
PT develop agents for diagnosis and treatment of diseases associated
PT with disruption of chromosome II at q23
PS Disclosure; Page 43-54; 68pp; English.
CC In the course of the construction of a physical map of human
CC chromosome region 11q23, a region contg. the t(4;11) translocation
CC breakpoint was cloned. The cloned DNA encoded a protein homologous
CC to the trithorax gene prod. of Drosophila. The gene may be used for
CC the diagnosis and treatment of immunodeficiency states,
CC developmental abnormalities, inherited diseases or cancers, e.g.
CC acute lymphocytic leukaemia or acute myelomonocytic leukaemia.
CC See also R52972-7.
SQ Sequence 3969 AA;

DB 9; Score 150; Match 27.7%; QryMatch 1.6%; Pred. No. 4.94e-02;
Matches 13; Conservative 18; Mismatches 15; Indels 1; Gaps 1;

Db 848 rnkdkapeelskdrdadksvekdksrerdxerekenkresxkekrkk 894
Qy 1246 RRRDRD-EDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDR 1291

RESULT 5
ID P92275 standard; peptide; 765 AA.
AC P92275;
DT 27-Feb-1990 (first entry)
DE Human topoisomerase I cDNA
KW Scleroderma.
OS Homo sapiens (human).
PN WO8909222-A.
PD 05-OCT-1989.
PF 22-MAR-1989; U01116.
PR 23-MAR-1988; US-172159.
PA (BRIG) Brigham and Women's Hospital; (UYJO) John's Hopkins Univ.
PI Earnshaw WC, D'Arpa P;
DR WPI; 89-309500/42.
DR N-PSDB; N91475.
PT Cloned cDNA encoding eukaryotic topoisomerase I - useful for large scale
PT prodn. by recombinant methods
PS Claim 6; fig. 5; 28pp; English.
CC The cDNA of this can be spliced into DNA vectors and used to transform
CC hosts for high yield. This polypeptide (I) retains the ability to bind
CC autoantibodies, even though the prokaryotic host degrades transcribed (I)
CC into a spectrum of polypeptides. (I) may be used to classify patients
CC with immune rheumatic diseases.
SQ Sequence 765 AA;

DB 1; Score 144; Match 17.7%; QryMatch 1.5%; Pred. No. 1.21e-01;
Matches 11; Conservative 28; Mismatches 21; Indels 2; Gaps 2;

Db 17 rlndshkhkdkhkdrehrhkehkkekdrekskhsnseh-kdsekkhkekektkhkdgsse 75
Qy 1246 RRRDRDEDRDRDRERD-RDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASPRE 1304

Db 76 kh 77
Qy 1305 KR 1306

RESULT 6
ID R12844 standard; Protein; 265 AA.
AC R12844;
DT 18-SEP-1991 (first entry)
DE HTLV-1 protein expressed by antisense nucleotides 1589-2383.
KW human T-cell lymphotrophic virus-1; ORF; vaccine; antisense RNA.
OS Human T cell lymphotrophic virus.
PN US4999421-A.
PD 12-MAR-1991.
PF 27-JUN-1988; 211749.
PR 27-JUN-1988; US-211749.
PA (TRIT-) TRITON BIOSCIENCES.
PI Brunck TK, Larocca DJ, Monahan JJ;
DR WPI; 91-206841/28.
DR N-PSDB; Q12502.
PT Proteins encoded by ribonucleic acid anti-sense strand - to human
PT T-cell leukaemia virus 1 viral RNA useful as diagnostic agents,
PT vaccines and therapeutic agents
PS Claim 1; Column 16; 15pp; English.
CC This polypeptide is encoded by nucleotides 1589-2383 in the reverse
CC complement of the HTLV-I genome (complementary to nucleotides 7479-
CC 6685, 3'-5' of the HTLV-I sense mRNA). Peptide fragments
CC corresponding to at least 10 amino acids from this sequence are
CC disclosed. See also Q12499-Q12501 and Q12503.
SQ Sequence 265 AA;

DB 3; Score 144; Match 25.0%; QryMatch 1.5%; Pred. No. 1.21e-01;
Matches 20; Conservative 27; Mismatches 29; Indels 4; Gaps 4;

Db 122 pgekapprgethrdrqr-raeek-rkrkkerekeeekqtaeylkr-keeekarrrrraek 178
Qy 1232 PSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDR 1291

Db 179 kaadvarrkqeeqerrerkw 198
Qy 1292 YRDERDHR-ASPREKRQRFW 1310

RESULT 7
ID R27530 standard; Protein; 740 AA.
AC R27530;
DT 08-MAR-1993 (first entry)
DE Plasmodium falciparum bloodand liver stage ABRA antigen.
KW Recombinant poxvirus; antimalarial vaccine; malaria; immunise;
KW immunogen; ss.
OS Plasmodium falciparum.
PN WO9216616-A.
PD 01-OCT-1992.
PF 19-MAR-1992; U02207.
PR 20-MAR-1991; US-672183.
PR 18-MAR-1992; US-852305.
PA (VIRO-) VIROGENETICS CORP.
PI De TAISNE C, Paoletti E, Tine JA;
DR WPI; 92-349203/42.
DR N-PSDB; Q29187.
PT Recombinant pox:virus - contg. Plasmodium DNA, useful as
PT antimalarial vaccine
PS Example 2; Fig 3; 74pp; English.
CC This sequence is the Plasmodium falciparum blood and liver stage antigen

```
CC   ABRA. cDNA encoding it was cloned into vaccinia donor plasmids
CC   before being inserted into the vaccinia virus to be used in a
CC   vaccine to stimulate an antimalarial immunological response, or for
CC   in vitro prodn. of gene prods. for use as immunogens. As plasmodium
CC   genes are conserved among P. falciparum strains, they are widely
CC   effective in a vaccine.
SQ   Sequence   740 AA;

  DB  5;  Score     137;  Match 9.6%;  QryMatch 1.4%;  Pred. No. 3.37e-01;
  Matches     5;  Conservative   37;  Mismatches   6;  Indels   4;  Gaps   4;

Db    672 knke-eeekekekekeke-ekeke-ekekeke-ekekekekeekeeekkeke 719
          : :: :|:::::::|:::: ::::: ::::::: ::|: :|:: ::  : |::
Qy   1246 RRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERD 1297


RESULT    8
ID   R51053 standard; peptide; 123 AA.
AC   R51053;
DT   21-OCT-1994  (first entry)
DE   Epstein-Barr nuclear antigen EBNA-1 immunoreactive fragment.
KW   Epstein-Barr Virus; binding epitope; nuclear antigen; EBNA-1;
KW   immunoreactive; anti-EBV antibody.
OS   Epstein-barr virus.
PN   WO9406912-A.
PD   31-MAR-1994.
PF   13-SEP-1993; E02478.
PR   14-SEP-1992; EP-202797.
PA   (ALKU ) AKZO NV.
PI   Middeldorp JM.
DR   WPI; 94-118461/14.
PT   New Epstein-Barr virus peptide(s) and antibodies - used in
PT   diagnosis, immuno:cyto:chemistry, typing and therapy involving
PT   Epstein Barr virus-related disease
PS   Claim 1; Page 39; 56pp; English.
CC   Peptides comprising at least part of amino acid sequence R51053 and
CC   which retain immunoreactivity with antibodies to Epstein-Barr virus
CC   are claimed. The peptides, which correspond to a binding epitope
CC   highly conserved in the EBV family, are useful for the detection or
CC   production of anti-EBV antibodies and treatment of EBV-related
CC   diseases. See also R51054-R51058.
SQ   Sequence   123 AA;

  DB  9;  Score     133;  Match 31.5%;  QryMatch 1.4%;  Pred. No. 6.02e-01;
  Matches    17;  Conservative    8;  Mismatches  28;  Indels   1;  Gaps   1;

Db      7 rgrgrerarggsrerargrgrgrgekrprspssqssssgspprrppp-grrpff 59
          | | |:| |   |:| | | | |   |:|    :        :|::|  :|  |
Qy   1256 RDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASPREKRQRF 1309


RESULT    9
ID   R13319 standard; Protein; 1023 AA.
AC   R13319;
DT   22-OCT-1991  (first entry)
DE   Partial Human Natural Killer receptor.
KW   NK; cytotoxic drugs; tumour cell.
OS   Homo sapiens.
PN   US7535206-A.
PD   09-JUL-1991.
PF   08-JUN-1990; 143578.
PR   08-JUN-1990; US-535206.
PA   (USSH ) NAT INST OF HEALTH.
PI   Ortaldo J, Young H, Anderson S;
```



```
DR   WPI; 91-245694/33.
DR   N-PSDB; Q13114.
PT   DNA encoding a natural killer cell receptor - used to develop
PT   prods. for the immuno-detection and immuno-therapy of tumours
PS   Disclosure; Fig 1; 30pp; English.
CC   Overlapping clones, which make up the cDNA sequence from which this
CC   sequence was deduced, were isolated from a cDNA library prepared
CC   from human NK cells purified from human peripheral blood. The
CC   purified protein can mediate the cytolytic activity of mammalian
CC   cells. It specifically distinguishes tumour cells making it a
CC   candidate for the development of products for the immunodetection
CC   and immunotherapy of tumours. See also Q13115.
SQ   Sequence   1023 AA;

  DB  3;  Score     133;  Match 26.3%;  QryMatch 1.4%;  Pred. No. 6.02e-01;
  Matches    54;  Conservative   43;  Mismatches  92;  Indels  16;  Gaps  15;

Db    574 rsrtaksssshsrsrsksrssssksghrk-rasks-prktasqlse--nkpvk-teplrat 628
          |  |:|   |  ::| || ||| :| :  |:||| ||  |: |:   : |     : ::
Qy   1115 RLSTPSPKCSPIHPR-KSSSSSGGGDKSSRTSKSAPRPCAPSLTTITEEPSSWHSSAHSV 1173

Db    629 maqnenvvvqpvvaenipviplsdspppsrwkpgekpwkpsyeriqemkaktthllpiqs 688
           : :::|||| |  :      |||:: :|  |  |:   :   |   |  :|  : ::
Qy   1174 QSSMQSIVVQPEVVVETTTYNGSDSAS-GRSTPT-KSSHGG-A-ITTTKVTATANIKVEV 1229

Db    689 tyslaniketgssssy-hkreknsesdqstyskysdrssesspsrsrsrsssrsrsysrsyt 747
            : :: |   |   | ::|::: : |:    :  ||  :   | | |  | |   ||
Qy   1230 V-TPSDRKSRRSYHYYDRRRDRDEDRDRDR-ERDRDRDRDRD-RDRDRD-RDRDRERSRE 1285

Db    748 rsrslasshsrsrspssrshsrnky 772
          | |       | : :|:|   | ::
Qy   1286 RDRRDRYRDERDHRASPR-EKRQRF 1309


RESULT   10
ID   P60624 standard; Protein; 183 AA.
AC   P60624;
DT   13-AUG-1991  (first entry)
DE   Sequence B encoded by a portion of SA307.
KW   Antibiotic; biotin binding affinity; fusion protein.
OS   Streptomyces.
PN   WO8602077-A.
PD   10-APR-1986.
PF   01-OCT-1985; 001901.
PR   02-OCT-1984; US-656873.
PA   (MEAD/) MEADE H M.
PI   Meade HM, Garwin JL, Biogen NV;
DR   WPI; 86-106643/16.
DR   N-PSDB; N60626.
PT   DNA sequences and hybrid DNA sequences - encoding
PT   streptavidin-like polypeptide, also joined to another protein,
PT   e.g. tissue plasminogen activator
PS   Disclosure; Fig. 2; 54pp; English.
CC   The inventors claim the DNA sequence in SA307 which codes for a
CC   steptavidin-like polypeptide (see N60626), and the polypeptide
CC   encoded by it (P60625). They also claim hybrid SQs comprising N60626
CC   and a second sequence coding for another protein, polypeptide,
CC   peptide or AA (pref. tissue plasminogen activator (TPA)).
SQ   Sequence   183 AA;

  DB  3;  Score     129;  Match 38.6%;  QryMatch 1.4%;  Pred. No. 1.07e+00;
  Matches    22;  Conservative    8;  Mismatches  25;  Indels   2;  Gaps   2;

Db     37 rgrhhrhlvqparldlhrdrgrrrrpdrnlrvgrrqrreplrpdrslrqrpghrrqr 93
```

```
          | | : :  :  | |  ||| | |  ||: |   |:||:  | :|  |  |  :|||
Qy   1254 RDRDRERDRDRDR-DRDRDRDRDRDRDRE-RSRERDRRDRYRDERDHRASPREKRQR 1308


RESULT   11
ID   P94776 standard; protein; 751 AA.
AC   P94776;
DT   05-JUL-1990  (first entry)
DE   Novel amyloid precursor protein (NAP3).
KW   SPAP; amyloid precursor protein; dysbolism; INS76; NAP3; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Active-site     289..345
FT   /*tag=  a
FT   /*label=INS57
PN   EP-304013-A.
PD   22-FEB-1989.
PF   16-AUG-1988; 113283.
PR   15-AUG-1987; JP-203298.
PR   21-AUG-1987; JP-207995.
PR   18-NOV-1987; JP-291404.
PR   11-DEC-1987; JP-313228.
PR   05-FEB-1988; JP-025260.
PR   10-FEB-1988; JP-029366.
PR   19-FEB-1988; JP-037905.
PR   25-MAY-1988; JP-125660.
PA   (ASAH) Asahi Kasei Kogyo.
PI   Kitaguchi N, Takahashi Y, Tokushima Y, Itoh H;
DR   WPI; 89-055458/08.
DR   N-PSDB; N91050.
PT   Human senile plague amyloid precursor protein and DNA -
PT   used for study and diagnosis of dysbolism in the central nervous
PT   system.
PS   Disclosure;  108pp; English.
CC   AA. sequence from 289 to 345 encodes INS57 which is claimed as a senile
CC   plaque amyloid precursor (SPAP), useful for diagnosis of dysbolism in the
CC   CNS such as senile dementia.
SQ   Sequence   751 AA;

  DB  1;  Score     129;  Match 23.5%;  QryMatch 1.4%;  Pred. No. 1.07e+00;
  Matches    23;  Conservative   36;  Mismatches  32;  Indels   7;  Gaps   7;

Db    342 gsaipttaastpdavdkyletpgdeneh-a-hf-qkakerleakhrermsqvmrewe-ea 397
          |:|| ||  ::   :   : ||:| : : : |: :: ::| | : |:|  :  |::: :
Qy   1211 GGAITTTKVTATANIKVEVVTPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDR 1270

Db    398 erqakhrermsqvmreweeaer-qesleqeaa-nerqq 433
          :|: : |:|  :  ||::  :| ::  :: |:  |::|
Qy   1271 DRD-RDRDRDRERSRERDRRDRYRDERDHRASPREKRQ 1307


RESULT   12
ID   R32082 standard; Protein; 179 AA.
AC   R32082;
DT   07-JUN-1993  (first entry)
DE   Plasmid pSGHEC-1414 peptide.
KW   Plasmid; fragment; fused; polypeptide; antigen; hepatitis C; serum;
KW   diagnosis.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   Peptide         106..179
PN   J05001099-A.
PD   08-JAN-1993.
PF   25-JUN-1991; 153031.
```



```
PR   25-JUN-1991; JP-153031.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
DR   WPI; 93-049639/06.
DR   N-PSDB; Q36418.
PT   New fused polypeptide antigens contg. hepatitis C related antigen
PT   - useful for exact serum diagnosis of hepatic C infections
PS   Disclosure; Page 23-24; 32pp; Japanese.
CC   The sequences given in R32078-84 are encoded by plasmid fragments
CC   which were used in the scope of the invention to produce fused
CC   polypeptide antigens containing hepatitis C related antigen.
CC   These polypetide antigens can be used for an exact serum diagnosis
CC   of a patient infected with type C hepatitis virus.
SQ   Sequence   179 AA;

  DB  6;  Score     129;  Match 26.1%;  QryMatch 1.4%;  Pred. No. 1.07e+00;
  Matches    18;  Conservative   23;  Mismatches  24;  Indels   4;  Gaps   4;

Db    111 dqiktkdrtqqrktkrstnrrrsknekkkkkefreqdqiktkdrtqqrktkrstnrrrsk 170
          : :   ||   |:: :  :||| ::| : : : ||:|: : :|| ::|   |  :| ||:
Qy   1228 EVVTPSDRKS-RRSYHYYDRRRDRDEDRDR-D-RERDRDRDRDRDRDRDRDRDRERSR 1284

Db    171 nekkkkkef 179
            |: ::  :
Qy   1285 -ERDRRDRY 1292


RESULT   13
ID   R14216 standard; Protein; 179 AA.
AC   R14216;
DT   07-JAN-1992  (first entry)
DE   HCV detecting peptide component (8).
KW   Diagnosis; immunoassay; antigen; hepatitis C virus.
PN   WO9114705-A.
PD   03-OCT-1991.
PF   22-MAR-1991; J00376.
PR   28-MAR-1990; JP-080520.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PI   Morita K, Mori H, Hasegawa M, Yokoo Y, Sato M, Sekine S;
PI   Sugimoto S;
DR   WPI; 91-310537/42.
DR   N-PSDB; Q14066.
PT   Fused antigen polypeptide - derived from hepatitis C virus and
PT   polypeptide(s) pref. from dog salmon growth hormone, fused via
PT   methionine, used for detecting hepatitis C
PS   Disclosure; Page 74-76; 109pp; Japanese.
CC   The peptides encoded by the sequences of Q14059-69 and Q14780 are
CC   disclosed to illustrate the invention which relates to fused antigen
CC   polypeptides.  The polypeptide consists of at least one polypeptide
CC   derived from or related to antigenic polypeptides of hepatitis C virus,
CC   fused to another polypeptide (pref. dog salmon growth hormone
CC   polypeptide) via methionine or other linking gps.
CC   The fused antigen polypeptide can be used in the selective and
CC   accurate diagnosis of hepatitis C by immunoassay of antibodies to HCV.
SQ   Sequence   179 AA;

  DB  3;  Score     129;  Match 26.1%;  QryMatch 1.4%;  Pred. No. 1.07e+00;
  Matches    18;  Conservative   23;  Mismatches  24;  Indels   4;  Gaps   4;

Db    111 dqiktkdrtqqrktkrstnrrrsknekkkkkefreqdqiktkdrtqqrktkrstnrrrsk 170
          : :   ||   |:: :  :||| ::| : : : ||:|: : :|| ::|   |  :| ||:
Qy   1228 EVVTPSDRKS-RRSYHYYDRRRDRDEDRDR-D-RERDRDRDRDRDRDRDRDRDRERSR 1284

Db    171 nekkkkkef 179
            |: ::  :
```

```
Qy   1285 -ERDRRDRY 1292


RESULT   14
ID   R29580 standard; Protein; 657 AA.
AC   R29580;
DT   22-APR-1993  (first entry)
DE   FMR-1 gene product.
KW   Fragile X disease; sex chromosome; X chromosome; X linked syndrome;
KW   X linked retardation; X linked manic depression; TKCR;
KW   Martin-Bell syndrome; CA polymorphisms; PCR analysis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1..3766
FT   /*tag= a
PN   WO9220825-A.
PD   26-NOV-1992.
PF   22-MAY-1992; U04447.
PR   24-MAY-1991; US-705490.
PR   29-AUG-1991; US-751891.
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PA   (UYEM-) UNIV EMORY SCHOOL.
PA   MEDICINE.
PI   Caskey CT, Nelson DL, Oostra BA, Pieretti M, Warren ST;
DR   WPI; 92-415801/50.
DR   N-PSDB; Q31890.
PT   Gene sequence, related probes and cosmid(s) - useful in
PT   diagnosing fragile X syndrome, X-linked metal retardation, manic
PT   depression, and Martin Bell syndrome
PS   Claim 14; Page 54; 75pp; English.
CC   This sequence represents the FMR-1 gene product. It was isolated
CC   as detailed in Q31890. The level of expression of the FMR-1 gene
CC   may be determined by determining the level of this protein. This is
CC   useful when trying to detect fragile X syndrome, X-linked retardation,
CC   X-linked manic depression, TKCR and Martin-Bell syndrome.
SQ   Sequence   657 AA;

  DB  6;  Score     127;  Match 44.2%;  QryMatch 1.3%;  Pred. No. 1.42e+00;
  Matches    23;  Conservative    3;  Mismatches  25;  Indels   1;  Gaps   1;

Db      5 rargraaarrrrrprrrrrrrrrrrrrrrrrrrrrrrlglerpqpts-rgr 55
          | | |   | | | | | | | | | | | | | | | | | |   || : :| | :
Qy   1254 RDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASPREK 1305


RESULT   15
ID   R31961 standard; Protein; 1141 AA.
AC   R31961;
DT   10-JUN-1993  (first entry)
DE   Human cardiac cGI PDE.
KW   cGMP; cAMP; phosphodiesterase; myocardial; positive inotropic agent;
KW   milrinone; amrinone; imazodan.
OS   Homo sapiens.
PN   US7801167-A.
PD   15-DEC-1992.
PF   03-DEC-1991; 801167.
PR   03-DEC-1991; US-801167.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Manganiello VC;
DR   WPI; 93-067438/08.
DR   N-PSDB; Q36780.
PT   New myocardial cGMP inhibited cAMP phosphodiesterase - useful as
PT   immunogens by being joined to polypeptide(s) e.g. haemocyanin,
PT   BSA, etc.
```



```
PS   Disclosure; Fig 1; 70pp; English.
CC   Mixed oligodeoxynucleotide probes (P5 and P8) based on the partial
CC   amino acid sequences of two platelet phosphodiesterase peptides PDE5
CC   and PDE8 were synthesised, as was an oligonucleotide based on a
CC   sequence from PDE P1A.  The oligonucleotides were used to screen a
CC   human heart lambda ZAP II cDNA library and positive colonies
CC   purified by four successive screening.  Clone n.13.2 contains the
CC   entire cGMP- inhibited PDE cDNA coding sequence.  Specific
CC   inhibition of myocardial cGMP inhibited cAMP phosphodiesterases is a
CC   primary mechanism of action for a number of positive inotropic
CC   agents, e.g. milrinone, amrinone, imazodan, etc.  The protein may
CC   be bound to a solid support or joined to other polypeptides to be used
CC   as immunogens, e.g. keyhole limpet haemocyanin, BSA, tetanus toxoid etc.
CC   The protein (or fragments) may be used to raise monoclonal antibodies
CC   and hybridomas.  See also R31962.
SQ   Sequence   1141 AA;

  DB  6;  Score     124;  Match 30.2%;  QryMatch 1.3%;  Pred. No. 2.17e+00;
  Matches    35;  Conservative   27;  Mismatches  49;  Indels   5;  Gaps   5;

Db    240 laylagvlgillaryveqilpqsaeaaprehlgsqliagtkedipvfkrrrrsssvvsae 299
          : :|: : |:|:   |  ||  :||  | ||   :| : :    |:   |: ||| :::
Qy   1082 IVFLGLIDGLLFFPIVLSILGPAAEVRPIEH-PERLSTPSPKCSPI-HPRKSSSSSGGGD 1139

Db    300 msgcsskshrrtslpcipreqlmghsewdhkrgprgsqssgtsitvdiavmgeata 355
          :|: :||| |  | : |  | | |  :::: ||| || |: |: |:|:
Qy   1140 KSSRTSKSAPRPCAPSLTTIT-EEPSSW-HS-SAHSVQSSMQSIVVQPEVVVETTT 1192

Search completed: Wed Jan 17 17:16:47 1996
Job time : 90 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:09:43 1996; MasPar time 28.57 Seconds
703.808 Million cell updates/sec
Tabular output not generated.

Title: >US-08-319-745-4
Description: (1:1311) from US08319745.pep
Perfect Score: 9491
Sequence: 1 MVAPDSEAPSNPRITAAHES.....YRDERDHRASPREKRQRFWT 1311

Scoring table: PAM 150
Gap 11

Searched: 43470 seqs, 15335248 residues

Database: swiss-prot31
1 part1
2 part2
3 part3
4 part4
5 part5
6 part6
7 part7
8 part8

Statistics: Mean 57.597; Variance 137.878; scale 0.418

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4527	47.7	1286	5	PATC_DROME	MEMBRANE PROTEIN PATC	0.00e+00
2	342	3.6	382	6	RDP_HUMAN	RD PROTEIN.	7.34e-31
3	339	3.6	1106	7	STC_DROME	SHUTTLE CRAFT PROTEIN	2.10e-30
4	329	3.5	375	6	RDP_MOUSE	RD PROTEIN (WL623).	6.91e-29
5	295	3.1	471	6	RU17_XENLA	U1 SMALL NUCLEAR RIBO	8.42e-24
6	255	2.7	614	6	RU17_HUMAN	U1 SMALL NUCLEAR RIBO	5.48e-18
7	255	2.7	448	6	RU17_DROME	U1 SMALL NUCLEAR RIBO	5.48e-18
8	251	2.6	185	7	T2_MOUSE	OCTAPEPTIDE-REPEAT PR	2.03e-17
9	225	2.4	633	8	YLS3_CAEEL	HYPOTHETICAL 70.7 KD	8.78e-14
10	202	2.1	494	7	SR75_HUMAN	PRE-MRNA SPLICING FAC	1.13e-10

11	198	2.1	483	7	VE2_HPV14	E2 PROTEIN.	3.81e-10
12	194	2.0	244	8	YL53_CAEEL	HYPOTHETICAL 29.0 KD	1.27e-09
13	190	2.0	475	7	U2AF_MOUSE	SPLICING FACTOR U2AF	4.22e-09
14	190	2.0	475	7	U2AF_HUMAN	SPLICING FACTOR U2AF	4.22e-09
15	186	2.0	330	6	RLX3_STAAU	RLX PROTEIN.	1.39e-08
16	182	1.9	502	7	VE2_HPV25	E2 PROTEIN.	4.51e-08
17	182	1.9	197	7	TRSF_DROME	FEMALE-SPECIFIC TRANS	4.51e-08
18	181	1.9	221	7	SC35_CHICK	SPLICING FACTOR SC35	6.04e-08
19	181	1.9	221	7	SC35_HUMAN	SPLICING FACTOR SC35	6.04e-08
20	174	1.8	429	7	TO34_YEAST	TOM34 PROTEIN.	4.61e-07
21	173	1.8	233	8	YOD2_CAEEL	HYPOTHETICAL 26.8 KD	6.14e-07
22	171	1.8	1523	7	SON_HUMAN	SON PROTEIN (SON3).	1.09e-06
23	169	1.8	349	7	SR55_DROME	SERINE-ARGININE PROTE	1.92e-06
24	167	1.8	1549	7	TRHY_SHEEP	TRICHOHYALIN.	3.39e-06
25	167	1.8	695	8	XE7_HUMAN	PROTEIN XE7.	3.39e-06
26	167	1.8	252	8	VPHE_NPVAC	POLYHEDRAL ENVELOPE P	3.39e-06
27	166	1.7	493	7	VE2_HPV19	E2 PROTEIN.	4.49e-06
28	165	1.7	568	2	DISC_DROME	DISCONNECTED PROTEIN.	5.95e-06
29	165	1.7	164	8	X16_HUMAN	PRE-MRNA SPLICING FAC	5.95e-06
30	163	1.7	346	8	YNP1_CAEEL	HYPOTHETICAL 42.2 KD	1.04e-05
31	162	1.7	1061	6	RNE_ECOLI	RIBONUCLEASE E (EC 3.	1.38e-05
32	161	1.7	767	7	TOP1_MOUSE	DNA TOPOISOMERASE I (	1.82e-05
33	160	1.7	767	7	TOP1_CRIGR	DNA TOPOISOMERASE I (	2.40e-05
34	159	1.7	57	4	HSP1_DIDMA	SPERM PROTAMINE P1 (C	3.16e-05
35	159	1.7	3969	4	HRX_HUMAN	ZINC FINGER PROTEIN H	3.16e-05
36	158	1.7	503	8	YOW5_CAEEL	HYPOTHETICAL 57.1 KD	4.17e-05
37	155	1.6	1023	3	GLT_DROME	GLUTACTIN PRECURSOR.	9.47e-05
38	154	1.6	514	7	VE2_HPV5B	E2 PROTEIN.	1.24e-04
39	152	1.6	241	8	YO87_CAEEL	HYPOTHETICAL 28.5 KD	2.13e-04
40	152	1.6	240	7	U2AG_HUMAN	SPLICING FACTOR U2AF	2.13e-04
41	150	1.6	264	7	TRA2_DROME	TRANSFORMER-2 SEX-DET	3.65e-04
42	150	1.6	1407	7	TRHY_RABIT	TRICHOHYALIN.	3.65e-04
43	149	1.6	461	7	VE2_HPV09	E2 PROTEIN.	4.77e-04
44	147	1.5	1898	7	TRHY_HUMAN	TRICHOHYALIN.	8.11e-04
45	146	1.5	488	7	VE2_HPV49	E2 PROTEIN.	1.06e-03

ALIGNMENTS

RESULT 1
ID PATC_DROME STANDARD; PRT; 1286 AA.
AC P18502;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE MEMBRANE PROTEIN PATCHED.
GN PTC.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RM 90058658
RA HOOPER J.E., SCOTT M.P.;
RL CELL 59:751-765(1989).
RN [2]
RP SEQUENCE FROM N.A.
RM 90015164
RA NAKANO Y., GUERRERO I., HIDALGO A., TAYLOR A., WHITTLE J.R.S.,
RA INGHAM P.W.;
RL NATURE 341:508-513(1989).
CC -!- FUNCTION: SEGMENTATION POLARITY PROTEIN. EXACT FUNCTIONNOT
CC KNOWN. PTC PROBABLY PARTICIPATES IN CELL INTERACTIONS THAT
CC ESTABLISH PATTERN WITHIN THE SEGMENT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

```
CC   -!- SIMILARITY: TO C.ELEGANS SEX-DETERMINING TRANSFORMER PROTEIN 2.
DR   EMBL; M28418; DMMPP1.
DR   EMBL; M28999; DMMPP2.
DR   EMBL; X17558; DMPTCR.
DR   PIR; A33468; A33468.
DR   PIR; S06119; S06119.
DR   FLYBASE; FBGN0003892; TUF.
KW   TRANSMEMBRANE; DEVELOPMENTAL PROTEIN; SEGMENTATION POLARITY PROTEIN.
FT   TRANSMEM     77     92       POTENTIAL.
FT   TRANSMEM    428    448       POTENTIAL.
FT   TRANSMEM    466    486       POTENTIAL.
FT   TRANSMEM    493    513       POTENTIAL.
FT   TRANSMEM    533    553       POTENTIAL.
FT   TRANSMEM    563    583       POTENTIAL.
FT   TRANSMEM    678    699       POTENTIAL.
FT   TRANSMEM    932    952       POTENTIAL.
FT   TRANSMEM    956    998       POTENTIAL.
FT   TRANSMEM   1008   1028       POTENTIAL.
FT   TRANSMEM   1057   1077       POTENTIAL.
FT   TRANSMEM   1083   1103       POTENTIAL.
FT   CARBOHYD    142    142       POTENTIAL.
FT   CARBOHYD    298    298       POTENTIAL.
FT   CARBOHYD    335    335       POTENTIAL.
FT   CARBOHYD    388    388       POTENTIAL.
FT   CARBOHYD    807    807       POTENTIAL.
FT   CARBOHYD   1181   1181       POTENTIAL.
FT   CARBOHYD   1258   1258       POTENTIAL.
FT   CARBOHYD   1283   1283       POTENTIAL.
FT   CONFLICT    111    111       R -> G (IN REF. 2).
FT   CONFLICT    274    274       G -> A (IN REF. 2).
FT   CONFLICT    332    332       R -> A (IN REF. 2).
FT   CONFLICT    636    636       P -> A (IN REF. 2).
FT   CONFLICT    862    864       DVF -> ASSPTELLRANCIRNR (IN REF. 2).
FT   CONFLICT    866    866       Y -> N (IN REF. 2).
SQ   SEQUENCE   1286 AA;  142915 MW;  8981681 CN;

  DB  5;  Score    4527;  Match 53.6%;  QryMatch 47.7%;  Pred. No. 0.00e+00;
  Matches   637;  Conservative  257;  Mismatches 260;  Indels  34;  Gaps  23;

Db      7 prvpdthgd--vvdeklfsdlyirtswvdaqvaldqidkgkargsrtaiylrsvfqshle 64
          ||:  :|      : :  :||||||||||| :|| :::||:  |:||:: :|: :| :|
Qy     12 PRITAAHESPCATEARHSADLYIRTSWVDAALALSELEKGNIEGGRTSLWIRAWLQEQLF 71

Db     65 tlgssvqkhagkvlfvailvlstfcvglksaqihskvhqlwiqeggrleaelaytqktig 124
            ||  :|  |||||||||||||||||||||||:| |||:|||||||||| ||  ::|
Qy     72 ILGCFLQGDAGKVLFVAILVLSTFCVGLKSAQIHTRVDQLWVQEGGRLEAELKYTAQALG 131

Db    125 edesathqlliqtthdpnasvlhpqallahlevlvkatavkvhlydtewglrdmcnmpst 184
          | :|:|||||:|||: ||: |:||| ||| || |:  || | ||:|| || |:|:|  ||
Qy    132 EADSSTHQLVIQTAKDPDVSLLHPGALLEHLKVVHAATRVTVHMYDIEWRLKDLCYSPSI 191

Db    185 psfegiyyieqilrhlipcsiitpldcfwegsqllgpesavvipglnqrllwttlnpasv 244
          | |||   || |: ::|||:|||||||||||| |||||: :: :| |:::| || |||  |
Qy    192 PDFEGYHHIESIIDNVIPCAIITPLDCFWEGSKLLGPDYPIYVPHLKHKLQWTHLNPLEV 251

Db    245 mqymkqkmseekisfdfetveqymkraaigsgymekpclnplnpncpdtapnknstqppd 304
          :: :| |:   |: | : |:| |||||:| |:|| |||||:| :|:|| |||||:| : ||
Qy    252 VEEVK-KL---KFQFPLSTIEAYMKRAGITSAYMKKPCLDPTDPHCPATAPNKKSGHIPD 307

Db    305 vgailsggcygyaakhmhwpeelivggrkrnrsghlrkaqalqsvvqlmtekemydqwqd 364
          |:| || ||||:||  |||||:|||||  || :: |||||: ||:|||||  |:|||: | |
Qy    308 VAAELSHGCYGFAAAYMHWPEQLIVGGATRNSTSALRKARXLQTVVQLMGEREMYEYWAD 367

Db    365 nykvhhlgwtqekaaevlnawqrnfsreveqllrkqsriatnydiyvfssaalddilakf 424
```



```
          :||||::|| ||||| ||:||||:|: || :    : ::: | :| ||:::|:|||:||
Qy    368 HYKVHQIGWNQEKAAAVLDAWQRKFAAEVRKI-TTSGSVSSAYSFYPFSTSTLNDILGKF 426

Db    425 shpsalsivigvavtvlyafctllrwrdpvrgqssvgvagvllmcfstaaglglsallgi 484
          |  |  :|::|    ::|   ||::||||:|:|::||:|||||: :: |||||: |||||
Qy    427 SEVSLKNIILGYMFMLIYVAVTLIQWRDPIRSQAGVGIAGVLLLSITVAAGLGFCALLGI 486

Db    485 vfnaastqvvpflalglgvdhifmltaayae-s-n-rreq-tklilkkvgpsilfsa-cs 539
          |||:|||:|||||||||: :|:|:| :| | : :  ||: | |:||| | |:|::: |:
Qy    487 PFNASSTQIVPFLALGLGVQDMFLLTHTYVEQAGDVPREERTGLVLKKSGLSVLLASLCN 546

Db    540 tagsffaaafipvpalkvfclqaaivmcsnlaaallvfpamisldlrrrtagradifccc 599
            :|:|||::|:|||::||||||||:: ||:: ||||||||||||||||:|:|||::||
Qy    547 VM-AFLAAALLPIPAFRVFCLQAAILLLFNLGSILLVFPAMISLDLRRRSAARADLLCC- 604

Db    600 fpvwkeqpkvappvlplnnnngrgarhpkscnnnrvplpaqnplleqradipgsshslas 659
          : :    |  :|  || ::   | |:  |: : :|:   :: || : ::    :  | |
Qy    605 L-M----PE-SP--LPKKKIPER-AKTRKNDKTHRID-TTRQPLDPDVSENVTKTCCL-S 653

Db    660 fslatfafqhytpflmrswvkfltvmgflaalisslyastrlqdgldiidlvpkdsnehk 719
            ||:  |  :|:||:||: ||  :::::::| :::|: ::|:: ||||: |:|| :::||
Qy    654 VSLTKWAKNQYAPFIMRPAVKVTSMLALIAVILTSVWGATKVKDGLDLTDIVPENTDEHE 713

Db    720 fldaqtrlfgfysmyavtqgnfeyptqqqllrdyhdsfvrvphvikndngglpdfwlllf 779
          ||  | : ||||:|||||||||||||| | || :||| |||:|:::|||||||  ||| ||
Qy    714 FLSRQEKYFGFYNMYAVTQGNFEYPTNQKLLYEYHDQFVRIPNIIKNDNGGLTKFWLSLF 773

Db    780 sewlgnlqkifdeeyrdgrltkecwfpnassdailayklivqtghvdnpvdkelvlt-nr 838
            :|| :||  || |   | :| | |  ||| ::|||||:|||||||||:|| |: : :|
Qy    774 RDWLLDLQVAFDKEVASGCITQEYWCKNASDEGILAYKLMVQTGHVDNPIDKSLITAGHR 833

Db    839 lvnsdgiinqrafynylsawatndvfaygasqgklypeprqyfhqpneydlkipkslplv 898
          ||: ||||| :|||||||||||||| :|||||||:| |:|:: :| |::  | | || ||:
Qy    834 LVDKDGIINPKAFYNYLSAWATNDALAYGASQGNLKPQPQRWIHSPEDVHLEIKKSSPLI 893

Db    899 yaqmpfylhgltdtsqiktlighirdlsvkyegfglpnypsgipfifweqymtlrsslam 958
          |:|:|||| ||:|| |||||  :||| :||| |:|||:|||||:|||||: ||:|| :
Qy    894 YTQLPFYLSGLSDTXSIKTLIRSVRDLCLKYEAKGLPNFPSGIPFLFWEQYLYLRTSLLL 953

Db    959 ilacvllaalvlvsllllsvwaavlvilsvlasla-qifgamtllgiklsaipavilils 1017
            ||| | | :: | :|||: |||||| |: ||:|  |::| |:|||:||||:|||:|:|:
Qy    954 ALACALAAVFIAVMVLLLNAWAAVLVTLA-LATLVLQLLGVMALLGVKLSAMPAVLLVLA 1012

Db   1018 vgmmlcfnvlislgfmtsvgnrqrrvqlsmqmslgplvhgmltsgvavfmlstspfefvi 1077
            :|  : | | : |||:||:| ::||  |:::  |:|:||| |::::|  ||::|   ||
Qy   1013 IGRGVHFTVHLCLGFVTSIGCKRRRASLALESVLAPVVHGALAAALAASMLAASECGFVA 1072

Db   1078 rhfcwlllvvlcvgacnsllvfpillsmvgpeaelvplehpdristpsplpvrsskrsgk 1137
          | | :||| :: :|  ::|| |||:||::|| ||: |:||||:|||||       | :
Qy   1073 RLFLRLLLDIVFLGLIDGLLFFPIVLSILGPAAEVRPIEHPERLSTPSPKCSPIHPRKSS 1132

Db   1138 syvvqgsrssrgscqkshhhhhkdlndpslttiteepqswkssnssiq 1185
          |   | :||| |  ||  :       ||||||||| || ||  |:|
Qy   1133 SSSGGGDKSSRTS--KSAPRPC----APSLTTITEEPSSWHSSAHSVQ 1174


RESULT    2
ID   RDP_HUMAN      STANDARD;      PRT;   382 AA.
AC   P18615;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT   01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)
DE   RD PROTEIN.
GN   RD OR RDBP.
```

```
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   90126228
RA   SPEISER P.W., WHITE P.C.;
RL   DNA 8:745-751(1989).
RN   [2]
RP   SEQUENCE OF 56-382 FROM N.A.
RC   TISSUE=LIVER;
RM   90362680
RA   SUROWY C.S., HOGANSON G., GOSINK J., STRUNK K., SPRITZ R.A.;
RL   GENE 90:299-302(1990).
CC   -!- THE RD PROTEIN MIGHT NOT BIND RNA IN VIVO, POSSIBLY BECAUSE ITS
CC       RNP CONSENSUS SEQUENCE IS HIGHLY ATYPICAL.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: TO NUCLEAR RNA-BINDING PROTEINS.
CC   -!- SIMILARITY: 92% IDENTITY TO MOUSE RD PROTEIN.
DR   EMBL; X16105; HSRD.
DR   EMBL; M32274; HSMHRD1.
DR   EMBL; M32275; HSMHRD2.
DR   EMBL; M32276; HSMHRD3.
DR   EMBL; M33230; HSMHRD4.
DR   EMBL; M33231; HSMHRD5.
DR   PIR; A33640; A33640.
DR   MIM; 154040; 11TH EDITION.
KW   RNA-BINDING; RIBONUCLEOPROTEIN; REPEAT; NUCLEAR PROTEIN.
FT   DOMAIN      185    244       TANDEM REPEATS OF R-D.
FT   DOMAIN      251    339       RNP DOMAIN.
FT   CONFLICT     95     95       A -> K (IN REF. 2).
FT   CONFLICT    120    120       MISSING (IN REF. 2).
FT   CONFLICT    175    175       S -> M (IN REF. 2).
FT   CONFLICT    267    267       MISSING (IN REF. 2).
FT   CONFLICT    314    314       A -> V (IN REF. 2).
FT   CONFLICT    349    349       A -> S (IN REF. 2).
SQ   SEQUENCE   382 AA;  43384 MW;  642669 CN;

  DB  6;  Score     342;  Match 57.7%;  QryMatch 3.6%;  Pred. No. 7.34e-31;
  Matches    45;  Conservative   15;  Mismatches  16;  Indels   2;  Gaps   2;

Db    180 saspprsrsrdrshernrdrdrdrerdrdrdrdrdrerdrdrdrdrdrdrerdrdrerdr 239
          ::|  :||   :  :| |||| ||:|||:|||||||:|||||||||||||||| |:|:| |
Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

Db    240 drdregpfrrsdsfperr 257
          || |:   :|: :  |:|
Qy   1290 DRYRDERDHRASP-REKR 1306


RESULT    3
ID   STC_DROME      STANDARD;      PRT;  1106 AA.
AC   P40798;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   SHUTTLE CRAFT PROTEIN.
GN   STC.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RA   STROUMBAKIS N.D., LI Z., TOLIAS P.P.;
```



```
RL   SUBMITTED (APR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR   EMBL; U09306; DM09306.
DR   FLYBASE; FBGN0001978; STC.
KW   DNA-BINDING; NUCLEAR PROTEIN.
SQ   SEQUENCE   1106 AA;  123261 MW;  5710081 CN;

  DB  7;  Score     339;  Match 39.2%;  QryMatch 3.6%;  Pred. No. 2.10e-30;
  Matches    49;  Conservative   32;  Mismatches  42;  Indels   2;  Gaps   2;

Db    174 pefvpnfaklsleetpaaattngnst-asletainetrprtlraqepaergannqcsnhn 232
          || | : :  :  ::::: :|  :|: ::  |: : |    :: :  :    ::: | |
Qy   1184 PEVVVETTTYNGSDSASGRSTPTKSSHGGAITTTKVTATANIKVEVVTPSDRKSRRSYHY 1243

Db    233 yererererdrdrdrerdrdrdrdrdrdrdrdrdrdrdsrpgntrqqrrsdyrddredryer 292
          |:| |:|: ||||||||||||||||||||||||||    :  |: ||  |||:|: |
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERD-RRDRYRDERDHRASP 1302

Db    293 sdrrr 297
            ::|:
Qy   1303 REKRQ 1307


RESULT    4
ID   RDP_MOUSE      STANDARD;      PRT;   375 AA.
AC   P19426;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)
DE   RD PROTEIN (WL623).
GN   RD.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RM   88178091
RA   LEVI-STRAUSS M., CARROLL M.C., STEINMETZ M., MEO T.;
RL   SCIENCE 240:201-204(1988).
CC   -!- THE RD PROTEIN MIGHT NOT BIND RNA IN VIVO, POSSIBLY BECAUSE ITS
CC       RNP CONSENSUS SEQUENCE IS HIGHLY ATYPICAL.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- THIS PROTEIN IS ENCODED IN THE MHC III LOCUS.
CC   -!- SIMILARITY: TO NUCLEAR RNA-BINDING PROTEINS.
CC   -!- SIMILARITY: 92% IDENTITY TO HUMAN RD PROTEIN.
DR   EMBL; M21332; MMMHRDDK.
DR   PIR; A40112; A40112.
DR   HSSP; P19339; 1SXL.
KW   RNA-BINDING; RIBONUCLEOPROTEIN; REPEAT; NUCLEAR PROTEIN;
KW   MHC III.
FT   DOMAIN      184    254       TANDEM REPEATS OF R-D.
FT   DOMAIN      254    341       RNP DOMAIN.
SQ   SEQUENCE   375 AA;  42555 MW;  630579 CN;

  DB  6;  Score     329;  Match 57.5%;  QryMatch 3.5%;  Pred. No. 6.91e-29;
  Matches    46;  Conservative   15;  Mismatches  16;  Indels   3;  Gaps   3;

Db    179 saspprsrsrdrshdrsrdrdrdkerdrdrdrdrdrdrdkdkdrdrdrdrdrdkerdrdrdr 238
          ::|  :||   :  || |||| |::|||:||||||||||:|:||||||||:: |:||| |
Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

Db    239 drdr-erd-regpfrrsdsf 256
          || | ||| | :|  : : |
Qy   1290 DRYRDERDHRASP-REKRQ 1307
```

```
Qy   1290 DRYRDERDHRASPREKRQRF 1309


RESULT    5
ID   RU17_XENLA     STANDARD;      PRT;   471 AA.
AC   P09406;
DT   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KD.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   89210819
RA   ETZERODT M., VIGNALI R., SCHERLY D., MATTAJ I.W., CILIBERTO G.,
RA   PHILIPSON L.;
RL   EMBO J. 7:4311-4321(1988).
CC   -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1. IT BINDS
CC       STEM LOOP I OF U1 SNRNA.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
DR   EMBL; X12430; XLU170KR.
DR   PIR; S02016; S02016.
DR   PROSITE; PS00030; RNP_1.
KW   NUCLEAR PROTEIN; RIBONUCLEOPROTEIN; RNA-BINDING.
FT   DOMAIN      147    154       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      221    320       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      344    426       ARG/ASP/GLU-RICH (MIXED CHARGE).
SQ   SEQUENCE   471 AA;  57203 MW;  846584 CN;

  DB  6;  Score     295;  Match 50.0%;  QryMatch 3.1%;  Pred. No. 8.42e-24;
  Matches    43;  Conservative   19;  Mismatches  19;  Indels   5;  Gaps   5;

Db    339 ielkqepeeksrerdrerdrdrekgekdrdkdrdrdrdrrrshrdrdrekdrdrdrdrrr 398
          :|:   :: |||   :  || |:: | |||:||:||||| |  |||||::|||||||:| |
Qy   1227 VEVVTPSDRKSRRSYHYYDRRRDRDE-DRDRDRERDRDRDRD-RDRDRDRDRDRDRERSR 1284

Db    399 drdrdrerdkdhkrerdrgdrsekre 424
          :||| |:| :|  |:: |:   |||:
Qy   1285 ERDR-RDRYRDE-RDH-RASPREKRQ 1307


RESULT    6
ID   RU17_HUMAN     STANDARD;      PRT;   614 AA.
AC   P08621;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KD (SNRP70).
GN   SNRP70 OR RPU1 OR U1AP1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   87133480
RA   THEISSEN H., ETZERODT M., REUTER R., SCHNEIDER C., LOTTSPEICH F.,
RA   ARGOS P., LUHRMANN R., PHILIPSON L.;
RL   EMBO J. 5:3209-3217(1986).
RN   [2]
RP   SEQUENCE OF 553-608 FROM N.A.
RM   88096573
RA   SPRITZ R.A., STRUNK K., SUROWY C.S., HOCH S.O., BARTON D.E.,
RA   FRANCKE U.;
```



```
RL   NUCLEIC ACIDS RES. 15:10373-10391(1987).
CC   -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1. IT BINDS
CC       STEM LOOP I OF U1 SNRNA.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
DR   EMBL; X04654; HSRNP70K.
DR   EMBL; X07403; HSU170SB.
DR   PIR; A25707; A25707.
DR   MIM; 180740; 11TH EDITION.
DR   PROSITE; PS00030; RNP_1.
KW   NUCLEAR PROTEIN; RIBONUCLEOPROTEIN; RNA-BINDING.
FT   DOMAIN      321    328       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      408    487       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      521    570       ARG/ASP/GLU-RICH (MIXED CHARGE).
SQ   SEQUENCE   614 AA;  70081 MW;  1495006 CN;

  DB  6;  Score     255;  Match 47.6%;  QryMatch 2.7%;  Pred. No. 5.48e-18;
  Matches    30;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db    396 yderpgpsplphrdrdrdrerer-rersrerdkererrrsrsrdrrrrsrsrdkeerrrs 454
          || |        |||:|||:|:|:| |:| |:||::|:| ||| |||| | |  ::: |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD-ERDHRASP 1302

Db    455 rer 457
          ||:
Qy   1303 REK 1305


RESULT    7
ID   RU17_DROME     STANDARD;      PRT;   448 AA.
AC   P17133;
DT   01-AUG-1990 (REL. 15, CREATED)
DT   01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KD.
GN   SNRNP27D.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   90258833
RA   MANCEBO R., LO P.C.H., MOUNT S.M.;
RL   MOL. CELL. BIOL. 10:2492-2502(1990).
CC   -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1. IT BINDS
CC       STEM LOOP I OF U1 SNRNA.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
DR   EMBL; M31162; DMRNP70K.
DR   PIR; A36311; A36311.
DR   FLYBASE; FBGN0003458; SNRNP27D.
DR   PROSITE; PS00030; RNP_1.
KW   NUCLEAR PROTEIN; RIBONUCLEOPROTEIN; RNA-BINDING.
FT   DOMAIN      143    150       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      254    350       ARG/GLU-RICH (MIXED CHARGE).
SQ   SEQUENCE   448 AA;  52873 MW;  847591 CN;

  DB  6;  Score     255;  Match 39.2%;  QryMatch 2.7%;  Pred. No. 5.48e-18;
  Matches    29;  Conservative   18;  Mismatches  25;  Indels   2;  Gaps   2;

Db    295 srerydefdrrdrrdrererdrdrerekkkkrsksreressrer-rerkrerrdrergt- 352
          ||  |  :|||  || :|:|||:|:|:: : | : |:|:  |:| | | |:|||| |
Qy   1237 SRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDER 1296

Db    353 gsggdvkerkpdfr 366
            :  :|::  |:
Qy   1297 DHRASPREKRQRFW 1310
```

```
RESULT    8
ID   T2_MOUSE       STANDARD;      PRT;   185 AA.
AC   Q06666;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   OCTAPEPTIDE-REPEAT PROTEIN T2.
GN   T2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=MACROPHAGES;
RM   93092084
RA   DI CARLO M., MONTANA G., ROMANCINO D.P., MONTELEONE D.;
RL   J. SUBMICROSC. CYTOL. PATHOL. 24:467-472(1992).
DR   EMBL; X67863; MMT2.
KW   REPEAT.
FT   DOMAIN        2    159       20 X 8 AA APPROXIMATE TANDEM REPEATS.
SQ   SEQUENCE   185 AA;  22805 MW;  88702 CN;

  DB  7;  Score     251;  Match 41.8%;  QryMatch 2.6%;  Pred. No. 2.03e-17;
  Matches    28;  Conservative   18;  Mismatches  18;  Indels   3;  Gaps   3;

Db     18 erqrgreaerqrerqrgreaerqrgrergreaereaerqrgrerqrgrgeaerqrgrerg 77
          :|:| |: :|:|:|:| |: :|:| |:| |: :|: || | |:| | |   ||:: |
Qy   1245 DRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDR-RDRYRDERDH-RASP 1302

Db     78 reverhr 84
          ||  |:|
Qy   1303 RE-KRQR 1308


RESULT    9
ID   YLS3_CAEEL     STANDARD;      PRT;   633 AA.
AC   P34388;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 70.7 KD PROTEIN F09G8.3 IN CHROMOSOME III.
GN   F09G8.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RM   94150718
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RL   NATURE 368:32-38(1994).
DR   EMBL; L11247; CEF09G8.
DR   PIR; S44795; S44795.
```



```
DR   WORMPEP; F09G8.3; CE00138.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   633 AA;  70660 MW;  2049957 CN;

  DB  8;  Score     225;  Match 27.2%;  QryMatch 2.4%;  Pred. No. 8.78e-14;
  Matches    46;  Conservative   49;  Mismatches  71;  Indels   3;  Gaps   3;

Db     68 fpyyeqyltlsttvytlvvlvlfvafvtislflrvnlagslvtvfvllssylhlmewmyl 127
          | : |||| | |::   :   |   |::: ::|    |: |||:  | :  |:|:  | |
Qy    938 FLFWEQYLYLRTSLLLALACALAAVFIAVMVLLLNAWAAVLVTL-ALATLVLQLLGVMAL 996

Db    128 lgitvnvvsvinmamslgiaveffgqmlhgfynskkpkreerafaalvsngsttlsgifp 187
          ||: :: :: : : :::| :| |  ::  ||  |   ||  ||  || |  :  : | ::
Qy    997 LGVKLSAMPAVLLVLAIGRGVHFTVHLCLGFVTSIGCKRR-RASLALESVLAPVVHGALA 1055

Db    188 aimitagclsfadsrvlityfcnqlvgiglvcavhgvvymptllaifgs 236
          | : :|: |: ::   :   |   |: | ::  : |::: | :|:|:|:
Qy   1056 AAL-AASMLAASECGFVARLFLRLLLDIVFLGLIDGLLFFPIVLSILGP 1103


RESULT   10
ID   SR75_HUMAN     STANDARD;      PRT;   494 AA.
AC   Q08170;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   PRE-MRNA SPLICING FACTOR SRP75.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RM   93309435
RA   ZAHLER A.M., NEUGEBAUER K.M., STOLK J.A., ROTH M.B.;
RL   MOL. CELL. BIOL. 13:4023-4028(1993).
CC   -!- FUNCTION: A PROBABLE ROLE IN ALTERNATIVE SPLICE SITE SELECTION
CC       DURING PRE-MRNA SPLICING.
CC   -!- PTM: EXTENSIVELY PHOSPHORYLATED.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
DR   EMBL; L14076; HSSRP75A.
KW   MRNA PROCESSING; MRNA SPLICING; NUCLEAR PROTEIN; RNA-BINDING;
KW   PHOSPHORYLATION.
FT   DOMAIN        5      9       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN       35     43       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN       95    178       GLY-RICH.
FT   DOMAIN      179    494       ARG/SER-RICH.
SQ   SEQUENCE   494 AA;  56792 MW;  1196516 CN;

  DB  7;  Score     202;  Match 24.6%;  QryMatch 2.1%;  Pred. No. 1.13e-10;
  Matches    48;  Conservative   40;  Mismatches 103;  Indels   4;  Gaps   4;

Db    205 rsrsgsskssshsksr-srsrsgsrsrsksrsrsqsrsrskkeksrspskdksrsrshsag 263
          |  : |:| |   :| | | ||:  :|   |:| :|: :    : :      :| :||
Qy   1115 RLSTPSPKCSPIHPRKSSSSSGGGDKSSRTSKSAPRPCAPSLTTITEEPSSWHSSAHSVQ 1174

Db    264 ksrskskdqaeekiqnnd-nvgkpks-rspsrhkskskrsrsqerrveeekrgsveqgq 321
           |      |:|  ::    | : : | || :   | : : : :      : |   |  ::
Qy   1175 SSMQSIVVQPEVVVETTTYNGSDSASGRSTPTKSSHGGAITTTKVTATANIKVEVVTPSD 1234

Db    322 eqekslrqsrsrsrskagsrsrsrsrskskdkrksrkrsreesrsrsrsrskserskrrg 381
          :   :   | | :   | | | | : :|: : | | |:  | | ||| :  | | |
Qy   1235 RKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD 1294
```

```
Db    382 skrdskagsskkkkk 396
          || :|::  |: :
Qy   1295 E-RDHRASPREKRQR 1308


RESULT   11
ID   VE2_HPV14      STANDARD;      PRT;   483 AA.
AC   P36783;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   E2 PROTEIN.
GN   E2.
OS   HUMAN PAPILLOMAVIRUS TYPE 14.
OC   VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   94265501
RA   DELIUS H., HOFMANN B.;
RL   CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
CC   -!- FUNCTION: E2 IS A TRANSCRIPTIONAL TRANSACTIVATOR CAPABLE OF
CC       ACTIVATING A CONDITIONAL ENHANCER IN THE VIRAL LONG CONTROL
CC       REGION (LCR). E2 BINDS TO THE 5'-ACCGNNNCGGT-3' PALINDROMIC
CC       SEQUENCE.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER.
DR   EMBL; X74467; HPV14D.
DR   PIR; S36470; S36470.
DR   HSSP; P11299; 2BOP.
KW   EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
KW   TRANS-ACTING FACTOR.
SQ   SEQUENCE   483 AA;  54938 MW;  1194261 CN;

  DB  7;  Score     198;  Match 25.0%;  QryMatch 2.1%;  Pred. No. 3.81e-10;
  Matches    37;  Conservative   44;  Mismatches  54;  Indels  13;  Gaps  10;

Db    172 faddatrysktghwevkvnketvft-p-vtsstpp-espggqadsntsskttptatdsts 228
          ::::::    ::|  |  :::::    | |    |    ::: : :  :|::|::  :: :|:
Qy   1159 ITEEPSSWHSSAH-SVQSSMQSIVVQPEVVVETTTYNGSDSASGRSTPTKSSHGGAITTT 1217

Db    229 rlspadsrk-qs-qqantkgrr----ygrrpssrtrrttetrqrrrsrsksrsrsrsrsr 282
          ::::: : | :    :: |:||    | ||      |  : |:| | | : | | | | | |
Qy   1218 KVTATANIKVEVVTPSDRKSRRSYHYYDRRRDRDEDRDRD-RERDRDRDRDRDRDRDRDR 1276

Db    283 lrsrsrsqsserrsryrs-rs-rsrqke 308
          | | ||:  :|| |||  |  |:  :|
Qy   1277 DRDRERSRERDRRDRYRDERDHRASPRE 1304


RESULT   12
ID   YL53_CAEEL     STANDARD;      PRT;   244 AA.
AC   P34433;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 29.0 KD PROTEIN F44E2.3 IN CHROMOSOME III.
GN   F44E2.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RM   94150718
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
```



```
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RL   NATURE 368:32-38(1994).
CC   -!- SIMILARITY: TO DNAJ.
DR   EMBL; L23646; CEF44E2.
DR   PIR; S44822; S44822.
DR   WORMPEP; F44E2.3; CE00181.
KW   HYPOTHETICAL PROTEIN.
FT   DOMAIN        3     45       ARG/ASP/LYS-RICH.
FT   DOMAIN       79     90       PRO-RICH.
SQ   SEQUENCE   244 AA;  28994 MW;  260966 CN;

  DB  8;  Score     194;  Match 44.8%;  QryMatch 2.0%;  Pred. No. 1.27e-09;
  Matches    26;  Conservative    8;  Mismatches  22;  Indels   2;  Gaps   2;

Db      3 rrsrsrsrspkrdre-erkrredrdrdrerkrdr-kdrerkrrhrssssegsqaephq 58
          || |   |   |:|: :| |  ||||||:| ||| : ||| || |       :| |::
Qy   1247 RRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASPRE 1304


RESULT   13
ID   U2AF_MOUSE     STANDARD;      PRT;   475 AA.
AC   P26369;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   SPLICING FACTOR U2AF 65 KD SUBUNIT (U2 AUXILIARY FACTOR 65 KD
DE   SUBUNIT) (U2 SNRNP AUXILIARY FACTOR LARGE SUBUNIT).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL/AM;
RM   92279036
RA   SAILER A., MACDONALD N.J., WEISSMANN C.;
RL   NUCLEIC ACIDS RES. 20:2374-2374(1992).
CC   -!- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. BINDS TO THE
CC       POLYPYRIMIDINE TRACT OF INTRONS EARLY DURING SPLICEOSOME ASSEMBLY.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SUBUNIT: ASSOCIATES WITH A 35 KD PROTEIN.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RNP).
DR   EMBL; X64587; MMU2AF.
KW   NUCLEAR PROTEIN; RNA-BINDING; MRNA SPLICING; REPEAT.
FT   DOMAIN       25     63       ARG/LYS/SER-RICH.
FT   DOMAIN       64    182       INVOLVED IN BINDING U2AF-35.
FT   DOMAIN      151    156       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      195    202       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      260    265       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      300    307       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      377    382       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      429    436       RNA-BINDING (RNP1) (BY SIMILARITY).
SQ   SEQUENCE   475 AA;  53517 MW;  1206658 CN;

  DB  7;  Score     190;  Match 39.1%;  QryMatch 2.0%;  Pred. No. 4.22e-09;
  Matches    25;  Conservative   15;  Mismatches  19;  Indels   5;  Gaps   5;
```

```
Db      4 fdeferqlnenkqerdkenrhrkrshsrsrsrdrkr-rsrsrdr-rnrdqrsasrdrr-r 60
          :  ::|:  :  ::||:: | | | : | | ||| | | | |:| |:||:|   || | :
Qy   1241 YHYYDRR-RDRDEDRDRD-RERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDH 1298

Db     61 rskp 64
          |: |
Qy   1299 RASP 1302


RESULT   14
ID   U2AF_HUMAN     STANDARD;      PRT;   475 AA.
AC   P26368;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   SPLICING FACTOR U2AF 65 KD SUBUNIT (U2 AUXILIARY FACTOR 65 KD
DE   SUBUNIT) (U2 SNRNP AUXILIARY FACTOR LARGE SUBUNIT).
GN   U2AF65.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RM   92168111
RA   ZAMORE P.D., PATTON J.G., GREEN M.R.;
RL   NATURE 355:609-614(1992).
CC   -!- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. BINDS TO THE
CC       POLYPYRIMIDINE TRACT OF INTRONS EARLY DURING SPLICEOSOME ASSEMBLY.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SUBUNIT: ASSOCIATES WITH A 35 KD PROTEIN.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RNP).
DR   EMBL; X64044; HSU2AF.
DR   PIR; S20250; S20250.
DR   PIR; S26095; S26095.
DR   MIM; 191318; 11TH EDITION.
KW   NUCLEAR PROTEIN; RNA-BINDING; MRNA SPLICING; REPEAT.
FT   DOMAIN       25     63       ARG/LYS/SER-RICH.
FT   DOMAIN       64    182       INVOLVED IN BINDING U2AF-35.
FT   DOMAIN      151    156       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      195    202       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      260    265       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      300    307       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      377    382       RNA-BINDING (RNP2) (BY SIMILARITY).
FT   DOMAIN      429    436       RNA-BINDING (RNP1) (BY SIMILARITY).
SQ   SEQUENCE   475 AA;  53501 MW;  1208046 CN;

  DB  7;  Score     190;  Match 39.1%;  QryMatch 2.0%;  Pred. No. 4.22e-09;
  Matches    25;  Conservative   15;  Mismatches  19;  Indels   5;  Gaps   5;

Db      4 fdeferqlnenkqerdkenrhrkrshsrsrsrdrkr-rsrsrdr-rnrdqrsasrdrr-r 60
          :  ::|:  :  ::||:: | | | : | | ||| | | | |:| |:||:|   || | :
Qy   1241 YHYYDRR-RDRDEDRDRD-RERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDERDH 1298

Db     61 rskp 64
          |: |
Qy   1299 RASP 1302


RESULT   15
ID   RLX3_STAAU     STANDARD;      PRT;   330 AA.
AC   P14491;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
```



```
DE   RLX PROTEIN.
OS   STAPHYLOCOCCUS AUREUS.
OG   PLASMID PC223.
OC   PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RN154;
RA   DREHER J., MATZURA H.;
RL   SUBMITTED (SEP-1988) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: THIS PROTEIN IS PROBABLY REQUIRED FOR RELAXATION
CC       COMPLEX FORMATION AND PLASMID MOBILISATION BY CONJUGATIVE
CC       PLASMIDS.
DR   EMBL; X12831; SAPC223A.
DR   PIR; S28102; S28102.
KW   PLASMID.
SQ   SEQUENCE   330 AA;  38811 MW;  429700 CN;

  DB  6;  Score     186;  Match 31.8%;  QryMatch 2.0%;  Pred. No. 1.39e-08;
  Matches    21;  Conservative   24;  Mismatches  18;  Indels   3;  Gaps   3;

Db    233 fnrqneqreeqarqreleqar-rekikrdkerekewarfnrstqairqnrersereerer 291
          ::|: :: |:: |:|| :: | |::  ||:::|:::: | :|| :  |::| | ||: |
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDR-DRDRDRDRDRDR-ERSRERDRRDRYRDERDHRAS 1301

Db    292 erkare 297
           |  |:
Qy   1302 PREKRQ 1307

Search completed: Wed Jan 17 17:11:51 1996
Job time : 128 secs.
```

Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 17:12:10 1996; MasPar time 43.92 Seconds
711.918 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-4
Description: (1:1311) from US08319745.pep
Perfect Score: 9491
Sequence: 1 MVAPDSEAPSNPRITAAHES.....YRDERDHRASPREKRQRFWT 1311

Scoring table: PAM 150
Gap 11

Searched: 78488 seqs, 23849247 residues

Database: pir45
1 ann1
2 ann2
3 ann3
4 unann1
5 unann2
6 unann3
7 unann4
8 unann5
9 unann6
10 unann7
11 unrev1
12 unrev2

Statistics: Mean 55.100; Variance 168.754; scale 0.327

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	4527	47.7	1286	9	A33468	probable membrane pr	0.00e+00
2	4372	46.1	1299	9	S06119	membrane protein pat	0.00e+00
3	342	3.6	380	12	S36789	gene RD protein - hu	5.32e-24
4	342	3.6	382	9	A33640	class III histocompa	5.32e-24
5	342	3.6	325	9	JH0189	arginine/aspartate-r	5.32e-24
6	329	3.5	375	10	A40112	MHC-region RD-repeat	1.93e-22

7	300	3.2	1170	11	S52525	hypothetical protein	5.31e-19
8	295	3.1	471	9	S02016	U1 snRNP 70K protein	2.05e-18
9	260	2.7	378	10	S04336	U1 snRNP 70K protein	2.27e-14
10	255	2.7	301	10	S03048	U1 snRNP 70K protein	8.37e-14
11	255	2.7	437	12	S52484	ribonucleoprotein 68	8.37e-14
12	255	2.7	448	9	A36311	70K U1 small nuclear	8.37e-14
13	255	2.7	614	9	A25707	70K small nuclear ri	8.37e-14
14	255	2.7	437	12	S41225	70K protein - human	8.37e-14
15	239	2.5	303	8	S28147	U1 snRNP 70K protein	5.26e-12
16	225	2.4	633	9	S44795	F09G8.3 protein - Ca	1.85e-10
17	202	2.1	494	10	A48133	pre-mRNA splicing SR	5.60e-08
18	198	2.1	483	11	S36470	envelope protein - h	1.48e-07
19	197	2.1	783	10	A55817	cyclin-dependent kin	1.89e-07
20	194	2.0	244	9	S44822	F44E2.3 protein - Ca	3.89e-07
21	192	2.0	492	10	S19938	splicing factor U2AF	6.30e-07
22	192	2.0	492	12	S22646	splicing factor U2AF	6.30e-07
23	190	2.0	475	10	S20250	splicing factor U2AF	1.02e-06
24	188	2.0	484	9	A40988	54K arginine-rich nu	1.64e-06
25	186	2.0	330	6	S28102	rlx protein - Staphy	2.63e-06
26	182	1.9	197	9	A29648	female-specific tran	6.77e-06
27	182	1.9	502	6	S36494	envelope protein - h	6.77e-06
28	181	1.9	221	10	A42634	splicing factor SC35	8.57e-06
29	181	1.9	416	9	A48249	pre-mRNA splicing fa	8.57e-06
30	181	1.9	221	9	B42701	PR264 protein - chic	8.57e-06
31	181	1.9	221	9	S17327	PR264 protein - chic	8.57e-06
32	181	1.9	221	10	S17328	PR264 protein - huma	8.57e-06
33	181	1.9	221	10	A42701	PR264 protein - huma	8.57e-06
34	174	1.8	429	8	S45459	hypothetical protein	4.38e-05
35	173	1.8	233	9	S44882	ZC262.2 protein - Ca	5.51e-05
36	171	1.8	483	10	PN0099	Son3 protein - human	8.73e-05
37	169	1.8	350	9	A40459	nuclear phosphoprote	1.38e-04
38	169	1.8	350	9	S14620	RS55 protein - fruit	1.38e-04
39	167	1.8	1549	10	A40691	trichohyalin (EF han	2.18e-04
40	167	1.8	1549	10	S32633	trichohyalin - sheep	2.18e-04
41	167	1.8	252	6	C43679	ORF3 protein - Autog	2.18e-04
42	167	1.8	1203	9	S26650	DNA-binding protein	2.18e-04
43	166	1.7	493	6	S36488	envelope protein - h	2.73e-04
44	165	1.7	164	12	S14016	X16 protein - mouse	3.43e-04
45	165	1.7	568	9	S15008	gene disco protein -	3.43e-04

ALIGNMENTS

RESULT 1
ENTRY A33468 #type complete
TITLE probable membrane protein patched - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 27-Jan-1995
ACCESSIONS A33468
REFERENCE A33468
#authors Hooper, J.E.; Scott, M.P.
#journal Cell (1989) 59:751-765
#title The Drosophila patched gene encodes a putative membrane protein required for segmental patterning.
#cross-references MUID:90058658
#accession A33468
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-1286 ##label HOO
##cross-references GB:M28418; GB:M28999
##note nucleotide sequence is not given
KEYWORDS membrane protein

```
SUMMARY          #length 1286  #molecular-weight 142915  #checksum 1555

  DB  9;  Score    4527;  Match 53.6%;  QryMatch 47.7%;  Pred. No. 0.00e+00;
  Matches   637;  Conservative  257;  Mismatches 260;  Indels  34;  Gaps  23;

Db      7 prvpdthgd--vvdeklfsdlyirtswvdaqvaldqidkgkargsrtaiylrsvfqshle 64
Qy     12 PRITAAHESPCATEARHSADLYIRTSWVDAALALSELEKGNIEGGRTSLWIRAWLQEQLF 71

Db     65 tlgssvqkhagkvlfvailvlstfcvglksaqihskvhqlwiqeggrleaelaytqktig 124
Qy     72 ILGCFLQGDAGKVLFVAILVLSTFCVGLKSAQIHTRVDQLWVQEGGRLEAELKYTAQALG 131

Db    125 edesathqlliqtthdpnasvlhpqallahlevlvkatavkvhlydtewglrdmcnmpst 184
Qy    132 EADSSTHQLVIQTAKDPDVSLLHPGALLEHLKVVHAATRVTVHMYDIEWRLKDLCYSPSI 191

Db    185 psfegiyyieqilrhlipcsiitpldcfwegsqllgpesavvipglnqrllwttlnpasv 244
Qy    192 PDFEGYHHIESIIDNVIPCAIITPLDCFWEGSKLLGPDYPIYVPHLKHKLQWTHLNPLEV 251

Db    245 mqymkqkmseekisfdfetveqymkraaigsgymekpclnplnpncpdtapnknstqppd 304
Qy    252 VEEVK-KL---KFQFPLSTIEAYMKRAGITSAYMKKPCLDPTDPHCPATAPNKKSGHIPD 307

Db    305 vgailsggcygyaakhmhwpeelivggrkrnrsghlrkaqalqsvvqlmtekemydqwqd 364
Qy    308 VAAELSHGCYGFAAAYMHWPEQLIVGGATRNSTSALRKARXLQTVVQLMGEREMYEYWAD 367

Db    365 nykvhhlgwtqekaaevlnawqrnfsreveqllrkqsriatnydiyvfssaalddilakf 424
Qy    368 HYKVHQIGWNQEKAAAVLDAWQRKFAAEVRKI-TTSGSVSSAYSFYPFSTSTLNDILGKF 426

Db    425 shpsalsivigvavtvlyafctllrwrdpvrgqssvgvagvllmcfstaaglglsallgi 484
Qy    427 SEVSLKNIILGYMFMLIYVAVTLIQWRDPIRSQAGVGIAGVLLLSITVAAGLGFCALLGI 486

Db    485 vfnaastqvvpflalglgvdhifmltaayae-s-n-rreq-tklilkkvgpsilfsa-cs 539
Qy    487 PFNASSTQIVPFLALGLGVQDMFLLTHTYVEQAGDVPREERTGLVLKKSGLSVLLASLCN 546

Db    540 tagsffaaafipvpalkvfclqaaivmcsnlaaallvfpamisldlrrrtagradifccc 599
Qy    547 VM-AFLAAALLPIPAFRVFCLQAAILLLFNLGSILLVFPAMISLDLRRRSAARADLLCC- 604

Db    600 fpvwkeqpkvappvlplnnnngrgarhpkscnnnrvplpaqnplleqradipgsshslas 659
Qy    605 L-M----PE-SP--LPKKKIPER-AKTRKNDKTHRID-TTRQPLDPDVSENVTKTCCL-S 653

Db    660 fslatfafqhytpflmrswvkfltvmgflaalisslyastrlqdgldiidlvpkdsnehk 719
Qy    654 VSLTKWAKNQYAPFIMRPAVKVTSMLALIAVILTSVWGATKVKDGLDLTDIVPENTDEHE 713

Db    720 fldaqtrlfgfysmyavtqgnfeyptqqqllrdyhdsfvrvphviknqngglpdfwlllf 779
Qy    714 FLSRQEKYFGFYNMYAVTQGNFEYPTNQKLLYEYHDQFVRIPNIIKNDNGGLTKFWLSLF 773

Db    780 sewlgnlqkifdeeyrdgrltkecwfpnassdailayklivqtghvdnpvdkelvlt-nr 838
Qy    774 RDWLLDLQVAFDKEVASGCITQEYWCKNASDEGILAYKLMVQTGHVDNPIDKSLITAGHR 833

Db    839 lvnsdqiinqrafynylsawatndvfaygasqgklypeprqyfhqpneydlkipkslplv 898
Qy    834 LVDKDGIINPKAFYNYLSAWATNDALAYGASQGNLKPQPQRWIHSPEDVHLEIKKSSPLI 893
```



```
Db    899 yaqmpfylhgltdtsqiktlighirdlsvkyegfglpnypsgipfifweqymtlrsslam 958
Qy    894 YTQLPFYLSGLSDTXSIKTLIRSVRDLCLKYEAKGLPNFPSGIPFLFWEQYLYLRTSLLL 953

Db    959 ilacvllaalvlvsllllsvwaavlvilsvlasla-qifgamtllgiklsaipavililS 1017
Qy    954 ALACALAAVFIAVMVLLLNAWAAVLVTLA-LATLVLQLLGVMALLGVKLSAMPAVLLVLA 1012

Db   1018 vgmmlcfnvlislgfmtsvgnrqrrvqlsmqmslgplvhgmltsgvavfmlstspfefvi 1077
Qy   1013 IGRGVHFTVHLCLGFVTSIGCKRRRASLALESVLAPVVHGALAAAALAASMLAASECGFVA 1072

Db   1078 rhfcwlllvvlcvgacnsllvfpillsmvgpeaelvplehpdristpsplpvrsskrsgk 1137
Qy   1073 RLFLRLLLDIVFLGLIDGLLFFPIVLSILGPAAEVRPIEHPERLSTPSPKCSPIHPRKSS 1132

Db   1138 syvvqgsrssrgscqkshhhhhkdlndpslttiteepqswkssnssiq 1185
Qy   1133 SSSGGGDKSSRTS--KSAPRPC-----APSLTTITEEPSSWHSSAHSVQ 1174


RESULT    2
ENTRY           S06119     #type complete
TITLE           membrane protein patched - fruit fly (Drosophila
                  melanogaster)
ORGANISM        #formal_name Drosophila melanogaster
DATE            30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                  09-Sep-1994
ACCESSIONS      S06119
REFERENCE       S06119
  #authors      Nakano, Y.; Guerrero, I.; Hidalgo, A.; Taylor, A.; Whittle,
                  J.R.S.; Ingham, P.W.
  #journal      Nature (1989) 341:508-513
  #title        A protein with several possible membrane-spanning domains
                  encoded by the Drosophila segment polarity gene patched.
  #cross-references MUID:90015164
  #accession    S06119
    ##status       not compared with conceptual translation
    ##molecule_type DNA
    ##residues     1-1299 ##label NAK
GENETICS
  #gene         ptc
  #map_position 2 44D3-D4
KEYWORDS        glycoprotein; transmembrane protein
FEATURE
  74-92             #domain transmembrane #status predicted #label TM01\
  427-448           #domain transmembrane #status predicted #label TM02\
  456-503           #domain transmembrane #status predicted #label TM03\
  529-555           #domain transmembrane #status predicted #label TM04\
  557-585           #domain transmembrane #status predicted #label TM05\
  677-699           #domain transmembrane #status predicted #label TM06\
  967-1017          #domain transmembrane #status predicted #label TM07\
  1019-1047         #domain transmembrane #status predicted #label TM08\
  1061-1086         #domain transmembrane #status predicted #label TM09\
  1093-1121         #domain transmembrane #status predicted #label TM10\
  142,298,335,388,
  807,861,1194,1271 #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY         #length 1299  #molecular-weight 144091  #checksum 7740

  DB  9;  Score    4372;  Match 53.0%;  QryMatch 46.1%;  Pred. No. 0.00e+00;
  Matches   636;  Conservative  258;  Mismatches 260;  Indels  47;  Gaps  27;
```

```
Db      7 prvpdthgd--vvdeklfsdlyirtswvdaqvaldqidkgkargsrtaiylrsvfqshle 64
          ||:  :|      : :  :|||||||||| :|| :::||:  |:||:: :|: :| :|
Qy     12 PRITAAHESPCATEARHSADLYIRTSWVDAALALSELEKGNIEGGRTSLWIRAWLQEQLF 71

Db     65 tlgssvqkhagkvlfvailvlstfcvglksaqihskvhqlwiqegggleaelaytqktig 124
           ||  :|  ||||||||||||||||||||||||::| |||:|||| ||||| ||  ::|
Qy     72 ILGCFLQGDAGKVLFVAILVLSTFCVGLKSAQIHTRVDQLWVQEGGRLEAELKYTAQALG 131

Db    125 edesathqlliqtthdpnasvlhpqallahlevlvkatavkvhlydtewglrdmcnmpst 184
           | :|:||||:|||: ||: |:||| ||| || |:  || | ||:|| || |:|:|  ||
Qy    132 EADSSTHQLVIQTAKDPDVSLLHPGALLEHLKVVHAATRVTVHMYDIEWRLKDLCYSPSI 191

Db    185 psfegiyyieqilrhlipcsiitpldcfwegsqllgpesavvipglnqrllwttlnpasv 244
           | |||   || |: ::|||:||||||||||||| ||||: :: :| |:::| || |||  |
Qy    192 PDFEGYHHIESIIDNVIPCAIITPLDCFWEGSKLLGPDYPIYVPHLKHKLQWTHLNPLEV 251

Db    245 mqymkqkmseekisfdfetveqymkraaiasgymekpclnplnpncpdtapnknstqppd 304
           :: :| |:   |: | : |:| |||||:|:|:|| ||||:| :|:|| |||||:| : ||
Qy    252 VEEVK-KL---KFQFPLSTIEAYMKRAGITSAYMKKPCLDPTDPHCPATAPNKKSGHIPD 307

Db    305 vgailsggcygyaakhmhwpeelivggakrnrsghlrkaqalqsvvqlmtekemydqwqd 364
           |:| || ||||:||  |||||:|||||| || :: |||||: ||:||||| |:|||: | |
Qy    308 VAAELSHGCYGFAAAYMHWPEQLIVGGATRNSTSALRKARXLQTVVQLMGEREMYEYWAD 367

Db    365 nykvhhlgwtqekaaevlnawqrnfsreveqllrkqsriatnydiyvfssaalddilakf 424
           :||||::|| ||||| ||:|||:|: ||  :    : ::: | :| ||:::|:|||:||
Qy    368 HYKVHQIGWNQEKAAAVLDAWQRKFAAEVRKI-TTSGSVSSAYSFYPFSTSTLNDILGKF 426

Db    425 shpsalsivigvavtvlyafctllrwrdpvrgqssvgvagvllmcfstaaglglsallgi 484
           |  |  :|::|    ::|   ||::||||:|:|::||:|||||: :: |||||: |||||
Qy    427 SEVSLKNIILGYMFMLIYVAVTLIQWRDPIRSQAGVGIAGVLLLSITVAAGLGFCALLGI 486

Db    485 vfnaastqvvpflalglgvdhifmltaayae-s-n-rreq-tklilkkvgpsilfsa-cs 539
           |||:|||:|||||||||: :|:|| :| | : :  ||: | |:||| | |:|::: |:
Qy    487 PFNASSTQIVPFLALGLGVQDMFLLTHTYVEQAGDVPREERTGLVLKKSGLSVLLASLCN 546

Db    540 tagsffaaafipvpalkvfclqaaivmcsnlaaallvfpamisldlrrrtagradifccc 599
              :|:|||::|:||::||||||||::  ||:: |||||||||||||||:|:|||::||
Qy    547 VM-AFLAAALLPIPAFRVFCLQAAILLLFNLGSILLVFPAMISLDLRRRSAARADLLCC- 604

Db    600 fpvwkeqpkvappvlplnnnngrgarhpkscnnnrvalpaqnplleqradipgsshslas 659
           : :    |  :|  || ::   | |:  |: : :|:   :: ||  : ::    :  | |
Qy    605 L-M----PE-SP--LPKKKIPER-AKTRKNDKTHRID-TTRQPLDPDVSENVTKTCCL-S 653

Db    660 fslatfafqhytpflmrswvkfltvmgflaalisslyastrlqdgldiidlvpkdsnehk 719
           ||:  |  :|:||:||: ||  ::::::| :::|: ::|:: |||||: |:|| :::||
Qy    654 VSLTKWAKNQYAPFIMRPAVKVTSMLALIAVILTSVWGATKVKDGLDLTDIVPENTDEHE 713

Db    720 fldaqtrlfgfysmyavtqgnfeyptqqqllrdyhdsfvrvphviknqdngglpdfwlllf 779
           ||  | : ||||:|||||||||||||| | || :||| |||:|::|||||||  ||| ||
Qy    714 FLSRQEKYFGFYNMYAVTQGNFEYPTNQKLLYEYHDQFVRIPNIIKNDNGGLTKFWLSLF 773

Db    780 sewlgnlqkifdeeyrdgrltkecwfpnassdailayklivqtghvdnpvdkelvlt-nr 838
           :|| :||  || |   | :| | |  ||| ::|||||:||||||||:|| |: : :|
Qy    774 RDWLLDLQVAFDKEVASGCITQEYWCKNASDEGILAYKLMVQTGHVDNPIDKSLITAGHR 833

Db    839 lvnsdgiinqrafynylsawatnassptellrancirnrangasqgklypeprqyfhqpn 898
           ||: ||||| :|||||||||||| :     | |       |||||:| |:|:: :| |:
Qy    834 LVDKDGIINPKAFYNYLSAWATNDA-----L-A-Y------GASQGNLKPQPQRWIHSPE 880

Db    899 eydlkipkslplvyaqmpfylhgltdtsqiktlighirdlsvkyegfglpnypsgipfif 958
           :  | | || ||:|:|:|||| ||:||  ||||| :||| :|||: ||||:|||||||:|
Qy    881 DVHLEIKKSSPLIYTQLPFYLSGLSDTXSIKTLIRSVRDLCLKYEAKGLPNFPSGIPFLF 940
```



```
Db    959 weqymtlrsslamilacvllaalvlvsllllsvwaavlvilsvlasla-qifgamtllgi 1017
          ||||: ||:|| : ||| | | : | :|||: |||||| |: ||:|  |::| |:|||:
Qy    941 WEQYLYLRTSLLLALACALAAVFIAVMVLLLNAWAAVLVTLA-LATLVLQLLGVMALLGV 999

Db   1018 klsaipavililsvgmmlcfnvlislgfmtsvgnrqrrvqlsmqmslgplvhgmltsgva 1077
          ||||:|||:|:|:::|  : | | : |||:||:| ::||  |:::  |:|:||| |::::|
Qy   1000 KLSAMPAVLLVLAIGRGVHFTVHLCLGFVTSIGCKRRRASLALESVLAPVVHGALAAALA 1059

Db   1078 vfmlstspfefvirhfcwlllvvlcvgacnsllvfpillsmvgpeaelvplehpdristp 1137
             ||::|   || | | :||| :: :|  ::|| |||:||:|| ||: |:|||:|:|||
Qy   1060 ASMLAASECGFVARLFLRLLLDIVFLGLIDGLLFFPIVLSILGPAAEVRPIEHPERLSTP 1119

Db   1138 splpvrsskrsgksyvvqgsrssrgscqkshhhhhkdlndpslttiteepqswkssnssi 1197
          ||       | : |    | :||| |  ||  :       |||||||||| || ||  |:
Qy   1120 SPKCSPIHPRKSSSSSGGGDKSSRTS--KSAPRPC----APSLTTITEEPSSWHSSAHSV 1173

Db   1198 q 1198
          |
Qy   1174 Q 1174


RESULT    3
ENTRY           S36789     #type complete
TITLE           gene RD protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            09-Dec-1993 #sequence_revision 09-Dec-1993 #text_change
                  09-Dec-1993
ACCESSIONS      S36789
REFERENCE       S36789
   #authors     Cheng, J.; Macon, K.J.; Volanakis, J.E.
   #journal     Biochem. J. (1993) 294:589-593
   #title       cDNA cloning and characterization of the protein encoded by
                  RD, a gene located in the class III region of the human
                  major histocompatibility complex.
   #accession   S36789
      ##status       preliminary
      ##residues     1-380 ##label CHE
      ##cross-references EMBL:L03411
SUMMARY         #length 380  #molecular-weight 43239  #checksum 3281

  DB 12;  Score     342;  Match 57.7%;  QryMatch 3.6%;  Pred. No. 5.32e-24;
  Matches    45;  Conservative   15;  Mismatches  16;  Indels   2;  Gaps   2;

Db    179 sasprsrsrdrshernrdrdrdrerdrdrdrdrdrerdrdrdrdrdrdrerdrdrerdr 238
          ::|  :||   :  :| |||| ||:|||:|||||||:|||||||||||||| |:|:| |
Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

Db    239 drdregpfrrsdsfperr 256
          || |:   :|: :  |:|
Qy   1290 DRYRDERDHRASP-REKR 1306


RESULT    4
ENTRY           A33640     #type complete
TITLE           class III histocompatibility antigen RD - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            09-Mar-1990 #sequence_revision 18-Sep-1992 #text_change
                  18-Jun-1993
ACCESSIONS      A33640
REFERENCE       A33640
   #authors     Speiser, P.W.; White, P.C.
   #journal     DNA (1989) 8:745-751
   #title       Structure of the human RD gene: a highly conserved gene in
                  the class III region of the major histocompatibility
```

```
                  complex.
     #cross-references MUID:90126228
     #accession    A33640
        ##status        preliminary
        ##molecule_type DNA
        ##residues      1-382 ##label SPE
        ##cross-references GB:M32275; GB:M30115; GB:M32276; GB:M33230; GB:M33231
        ##note          the authors translated the codon AAG or residue 95 as
                         Ala
  SUMMARY          #length 382  #molecular-weight 43441  #checksum 8129

    DB  9;  Score     342;  Match 57.7%;  QryMatch 3.6%;  Pred. No. 5.32e-24;
    Matches    45;  Conservative   15;  Mismatches  16;  Indels   2;  Gaps   2;

  Db    180 saspprsrsrdrshernrdrdrdrerdrdrdrdrdrerdrdrdrdrdrdrerdrdrerdr 239
            ::|  :||   :  :| |||| ||:|||:|||||||:||||||||||||||| |:|:| |
  Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

  Db    240 drdregpfrrsdsfperr 257
            || |:   :|: :  |:|
  Qy   1290 DRYRDERDHRASP-REKR 1306


  RESULT    5
  ENTRY            JH0189     #type complete
  TITLE            arginine/aspartate-rich 37.3K protein - human
  ALTERNATE_NAMES  RD protein
  ORGANISM         #formal_name Homo sapiens #common_name man
  DATE             31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
                     24-Feb-1995
  ACCESSIONS       JH0189
  REFERENCE        JH0189
     #authors      Surowy, C.S.; Hoganson, G.; Gosink, J.; Strunk, K.; Spritz,
                     R.A.
     #journal      Gene (1990) 90:299-302
     #title        The human RD protein is closely related to nuclear
                     RNA-binding proteins and has been highly conserved.
     #cross-references MUID:90382680
     #accession    JH0189
        ##molecule_type mRNA
        ##residues      1-325 ##label SUR
        ##cross-references EMBL:X16105
  COMMENT     This protein consists of alternating basic and acidic amino acid,
                primarily arginine and asparitic acid, and contains a
                'ribonucleoprotein sequence domain'. This protein is closely
                related to nuclear RNA-binding proteins.
  CLASSIFICATION   #superfamily ribonucleoprotein repeat homology
  FEATURE
     195-282           #domain ribonucleoprotein #label RNP\
     208-267           #domain ribonucleoprotein repeat homology #label RRM3
  SUMMARY          #length 325  #molecular-weight 37276  #checksum 1492

    DB  9;  Score     342;  Match 57.7%;  QryMatch 3.6%;  Pred. No. 5.32e-24;
    Matches    45;  Conservative   15;  Mismatches  16;  Indels   2;  Gaps   2;

  Db    124 saspprsrsrdrshernrdrdrdrerdrdrdrdrdrerdrdrdrdrdrdrerdrdrerdr 183
            ::|  :||   :  :| |||| ||:|||:|||||||:||||||||||||||| |:|:| |
  Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

  Db    184 drdregpfrrsdsfperr 201
            || |:   :|: :  |:|
  Qy   1290 DRYRDERDHRASP-REKR 1306
```



```
  RESULT    6
  ENTRY            A40112     #type complete
  TITLE            MHC-region RD-repeat protein - mouse
  ORGANISM         #formal_name Mus musculus #common_name house mouse
  DATE             20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change
                     24-Feb-1995
  ACCESSIONS       A40112
  REFERENCE        A40112
     #authors      Levi-Strauss, M.; Carroll, M.C.; Steinmetz, M.; Meo, T.
     #journal      Science (1988) 240:201-204
     #title        A previously undetected MHC gene with an unusual periodic
                     structure.
     #cross-references MUID:88178091
     #accession    A40112
        ##status        preliminary
        ##molecule_type mRNA
        ##residues      1-375 ##label LEV
        ##cross-references GB:M21332
  CLASSIFICATION   #superfamily ribonucleoprotein repeat homology
  FEATURE
     267-326           #domain ribonucleoprotein repeat homology #label RRM3
  SUMMARY          #length 375  #molecular-weight 42555  #checksum 8746

    DB 10;  Score     329;  Match 57.5%;  QryMatch 3.5%;  Pred. No. 1.93e-22;
    Matches    46;  Conservative   15;  Mismatches  16;  Indels   3;  Gaps   3;

  Db    179 saspprsrsrdrshdrsrdrdrdkerdrdrdrdrdrdrdkdkdrdrdrdrdkerdrdrdr 238
            ::|  :||   :  || |||| |::|||:||||||||||:|:||||||||:: |:||| |
  Qy   1231 TPSDRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDR-R 1289

  Db    239 drdr-erd-regpfrrsdsf 256
            || | ||| | :|  : : |
  Qy   1290 DRYRDERDHRASPREKRQRF 1309


  RESULT    7
  ENTRY            S52525     #type complete
  TITLE            hypothetical protein - yeast (Saccharomyces cerevisiae)
  ORGANISM         #formal_name Saccharomyces cerevisiae
  DATE             08-May-1995 #sequence_revision 08-May-1995 #text_change
                     08-May-1995
  ACCESSIONS       S52525
  REFERENCE        S52519
     #authors      Badcock, K.; Churcher, C.
     #submission   submitted to the EMBL Data Library, February 1995
     #accession    S52525
        ##status        preliminary
        ##residues      1-1170 ##label BAD
        ##cross-references EMBL:Z48483
  SUMMARY          #length 1170  #molecular-weight 132644  #checksum 5191

    DB 11;  Score     300;  Match 28.7%;  QryMatch 3.2%;  Pred. No. 5.31e-19;
    Matches    52;  Conservative   64;  Mismatches  56;  Indels   9;  Gaps   8;

  Db    979 ldmfays-pfyiffvqyqtl-gpltlkligsaiiliffissvflqnirssfllalvvtmi 1036
            |  |: : || :|  ||  |  :| | | : |:  :|:   |:| | ::: |::| :: :
  Qy    929 LPNFPSGIPF-LFWEQYLYLRTSLLLAL-ACALAAVFIAVMVLLLNAWAAVLVTLALATL 986

  Db   1037 ivdigalmallgislnavslvnliicvglqvefcvhivrsftvvpsetkkdansrvlysl 1096
            :::: ::|||||: |:|:: | |:: :| || | ||:  :| | :   |:   |  | :|
  Qy    987 VLQLLGVMALLGVKLSAMPAVLLVLAIGRGVHFTVHLCLGF-VTSIGCKRRRAS--L-AL 1042

  Db   1097 ntigesvikgitltkfigvcvlafaqskifdvfyfrmwftliivaalhallflpallslf 1156
            :::  :|: | :|:  ::  :|| ::  :   :::|: : ::::: : :|||:| :||::
```

```
Qy   1043 ESVLAPVVHG-ALAAALAASMLAASECGFVARLFLRLLLDIVFLGLIDGLLFFPIVLSIL 1101

Db   1157 g 1157
          |
Qy   1102 G 1102
```

RESULT 8
ENTRY S02016 #type complete
TITLE U1 snRNP 70K protein - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 31-Jan-1995
ACCESSIONS S02016
REFERENCE S02016
#authors Etzerodt, M.; Vignali, R.; Ciliberto, G.; Scherly, D.; Mattaj, I.W.; Philipson, L.
#journal EMBO J. (1988) 7:4311-4321
#title Structure and expression of a Xenopus gene encoding an snRNP protein (U1 70K).
#cross-references MUID:89210819
#accession S02016
##molecule_type DNA
##residues 1-471 ##label ETZ
##cross-references EMBL:X12430
CLASSIFICATION #superfamily ribonucleoprotein repeat homology
FEATURE
104-174 #domain ribonucleoprotein repeat homology #label RRM
SUMMARY #length 471 #molecular-weight 57203 #checksum 6745

DB 9; Score 295; Match 50.0%; QryMatch 3.1%; Pred. No. 2.05e-18;
Matches 43; Conservative 19; Mismatches 19; Indels 5; Gaps 5;

```
Db    339 ielkqepeeksrerdrerdrdrekgekdrdkdrdrdrdrrrshrdrdrekdrdrdrdrrr 398
          :|:   :: |||   :  || |:: | |||:||:|||||| |  |||||::|||||||:| |
Qy   1227 VEVVTPSDRKSRRSYHYYDRRRDRDE-DRDRDRERDRDRDRD-RDRDRDRDRDRDRERSR 1284

Db    399 drdrdrerdkdhkrerdrgdrsekre 424
          :||| |:| :|  |:: |:   |||:
Qy   1285 ERDR-RDRYRDE-RDH-RASPREKRQ 1307
```

RESULT 9
ENTRY S04336 #type fragment
TITLE U1 snRNP 70K protein (long form) - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Feb-1990 #sequence_revision 30-Sep-1991 #text_change 31-Jan-1995
ACCESSIONS S04336
REFERENCE S04336
#authors Hornig, H.; Fischer, U.; Costas, M.; Rauh, A.; Luehrmann, R.
#journal Eur. J. Biochem. (1989) 182:45-50
#title Analysis of genomic clones of the murine U1RNA-associated 70-kDa protein reveals a high evolutionary conservation of the protein between human and mouse.
#cross-references MUID:89276388
#accession S04336
##molecule_type DNA
##residues 1-378 ##label HOR
##cross-references EMBL:X15769
##note the authors translated the codon GGC for residue 101 as Glu, ACT for residue 113 as Tyr, GAT for residue 202 as Glu, GAA for residue 280 as Asp, AGT for residue 350 as Met, and GAG for residues 368 and 374 as Asp

GENETICS
#introns 19/1; 40/3; 61/3; 89/1; 123/1; 152/2
CLASSIFICATION #superfamily ribonucleoprotein repeat homology
KEYWORDS alternative splicing
FEATURE
34-101 #domain ribonucleoprotein repeat homology #label RRM
SUMMARY #length 378 #checksum 2403

DB 10; Score 260; Match 49.2%; QryMatch 2.7%; Pred. No. 2.27e-14;
Matches 31; Conservative 13; Mismatches 17; Indels 2; Gaps 2;

```
Db    149 yderpgpsplphrdrdrdrererrersrsrdkererrrsrsrdrqrrsrsrd-kderrrs 207
          || |       |||:|||:|:| :| |:||::|:| | |||:|:|| | || :| |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDR-DRDRDRDRDRDRDRERSRERDRRDRYRDERDHRASP 1302

Db    208 rer 210
          ||:
Qy   1303 REK 1305
```

RESULT 10
ENTRY S03048 #type fragment
TITLE U1 snRNP 70K protein (clone RNP 8) - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 31-Jan-1995
ACCESSIONS S03048
REFERENCE S00674
#authors Spritz, R.A.; Strunk, K.; Surowy, C.S.; Hoch, S.O.; Barton, D.E.; Francke, U.
#journal Nucleic Acids Res. (1987) 15:10373-10391
#title The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding.
#cross-references MUID:88096573
#accession S03048
##molecule_type mRNA
##residues 1-301 ##label SPR
##cross-references EMBL:X06812
GENETICS
#map_position 19
CLASSIFICATION #superfamily ribonucleoprotein repeat homology
FEATURE
55-122 #domain ribonucleoprotein repeat homology #label RRM
SUMMARY #length 301 #checksum 9153

DB 10; Score 255; Match 47.6%; QryMatch 2.7%; Pred. No. 8.37e-14;
Matches 30; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

```
Db    170 yderpgpsplphrdrdrdrerer-rersrerdkererrrsrsrdrrrrsrsrdkeerrrs 228
          || |       |||:|||:|:| |:| |:|||::|:| ||| |||| | |  ::: |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD-ERDHRASP 1302

Db    229 rer 231
          ||:
Qy   1303 REK 1305
```

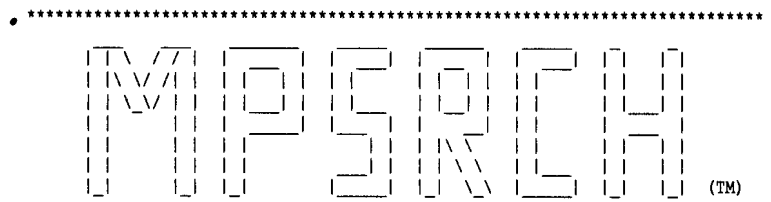
RESULT 11
ENTRY S52484 #type complete
TITLE ribonucleoprotein 68K (U1) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-May-1995 #sequence_revision 08-May-1995 #text_change 08-May-1995

```
  ACCESSIONS       S52484
  REFERENCE        S52484
    #authors       Northemann, W.; Berg, H.; Stahnke, G.; Walter, M.; Fenning,
                     S.
    #submission    submitted to the EMBL Data Library, February 1995
    #description   Identification of an inhibitory element within the human
                     68kDa (U1) ribonucleoprotein antigen.
    #accession     S52484
      ##status         preliminary
      ##residues       1-437 ##label NOR
      ##cross-references EMBL:X84841
  SUMMARY          #length 437  #molecular-weight 51556  #checksum 9783

   DB 12;  Score     255;  Match 47.6%;  QryMatch 2.7%;  Pred. No. 8.37e-14;
   Matches    30;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db    219 yderpgpsplphrdrdrdrerer-rersrerdkererrrsrsrdrrrrsrsrdkeerrrs 277
          || |        |||:|||:|:|:| |:| |:||::|:| ||| |||| | |  ::: |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD-ERDHRASP 1302

Db    278 rer 280
          ||:
Qy   1303 REK 1305


RESULT   12
ENTRY            A36311      #type complete
TITLE            70K U1 small nuclear ribonucleoprotein - fruit fly
                   (Drosophila melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             25-Jan-1991 #sequence_revision 25-Jan-1991 #text_change
                   31-Jan-1995
ACCESSIONS       A36311
REFERENCE        A36311
   #authors      Mancebo, R.; Lo, P.C.H.; Mount, S.M.
   #journal      Mol. Cell. Biol. (1990) 10:2492-2502
   #title        Structure and expression of the Drosophila melanogaster gene
                   for the U1 small nuclear ribonucleoprotein particle 70K
                   protein.
   #cross-references MUID:90258833
   #accession    A36311
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-448 ##label MAN
      ##cross-references GB:M31162
CLASSIFICATION   #superfamily ribonucleoprotein repeat homology
FEATURE
   103-170            #domain ribonucleoprotein repeat homology #label RRM
SUMMARY          #length 448  #molecular-weight 52873  #checksum 2804

  DB  9;  Score     255;  Match 39.2%;  QryMatch 2.7%;  Pred. No. 8.37e-14;
  Matches    29;  Conservative   18;  Mismatches  25;  Indels   2;  Gaps   2;

Db    295 srerydefdrrdrrdrererdrdrerekkkkrsksreressrer-rerkrerrdrergt- 352
          ||  |  :|||  || :|:|||:|:|:: : | : |:|:  |:| | | |:|||| |
Qy   1237 SRRSYHYYDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRERSRERDRRDRYRDER 1296

Db    353 gsggdvkerkpdfr 366
            :  :|::  |:
Qy   1297 DHRASPREKRQRFW 1310


RESULT   13
ENTRY            A25707      #type complete
```



```
TITLE            70K small nuclear ribonucleoprotein U1 - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change
                   28-Oct-1994
ACCESSIONS       A25707; A32289; B37108; A31274
REFERENCE        A25707
   #authors      Theissen, H.; Etzerodt, M.; Reuter, R.; Schneider, C.;
                   Lottspeich, F.; Argos, P.; Luehrmann, R.; Philipson, L.
   #journal      EMBO J. (1986) 5:3209-3217
   #title        Cloning of the human cDNA for the U1 RNA-associated 70K
                   protein.
   #cross-references MUID:87133480
   #accession    A25707
      ##molecule_type mRNA
      ##residues       1-614 ##label THE
      ##cross-references EMBL:X04654
REFERENCE        A32289
   #authors      Query, C.C.; Bentley, R.C.; Keene, J.D.
   #journal      Cell (1989) 57:89-101
   #title        A common RNA recognition motif identified within a defined U1
                   RNA binding domain of the 70K U1 snRNP protein.
   #cross-references MUID:89195226
   #accession    A32289
      ##molecule_type mRNA
      ##residues       178-614 ##label QUE
      ##cross-references GB:M22636
REFERENCE        A37108
   #authors      Spritz, R.A.; Strunk, K.; Surowy, C.S.; Mohrenweiser, H.W.
   #journal      Genomics (1990) 8:371-379
   #title        Human U1-70K ribonucleoprotein antigen gene: organization,
                   nucleotide sequence, and mapping to locus 19q13.3.
   #cross-references MUID:91065657
   #accession    B37108
      ##status         preliminary
      ##molecule_type DNA
      ##residues       178-614 ##label SPR
      ##cross-references GB:J04772
CLASSIFICATION   #superfamily ribonucleoprotein repeat homology
FEATURE
   281-348            #domain ribonucleoprotein repeat homology #label RRM
SUMMARY          #length 614  #molecular-weight 70081  #checksum 6564

  DB  9;  Score     255;  Match 47.6%;  QryMatch 2.7%;  Pred. No. 8.37e-14;
  Matches    30;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db    396 yderpgpsplphrdrdrdrerer-rersrerdkererrrsrsrdrrrrsrsrdkeerrrs 454
          || |        |||:|||:|:|:| |:| |:||::|:| ||| |||| | |  ::: |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD-ERDHRASP 1302

Db    455 rer 457
          ||:
Qy   1303 REK 1305


RESULT   14
ENTRY            S41225      #type complete
TITLE            70K protein - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             06-Oct-1994 #sequence_revision 06-Oct-1994 #text_change
                   06-Oct-1994
ACCESSIONS       S41225
REFERENCE        S41225
   #authors      Woppmann, A.; Will, C.L.; Kornstaedt, U.; Zuo, P.; Manley,
                   J.L.; Luehrmann, R.
```

```
    #journal      Nucleic Acids Res. (1993) 21:2815-2822
    #title        Identification of an snRNP-associated kinase activity that
                    phosphorylates arginine/serine rich domains typical of
                    splicing factors.
    #accession    S41225
      ##status        preliminary
      ##residues      1-437 ##label WOP
SUMMARY           #length 437  #molecular-weight 51556  #checksum 9783

  DB 12;  Score     255;  Match 47.6%;  QryMatch 2.7%;  Pred. No. 8.37e-14;
  Matches    30;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db    219 yderpgpsplphrdrdrdrerer-rersrerdkererrrsrsrdrrrrsrsrdkeerrrs 277
          || |        |||:|||:|:| |:| |:||::|:| ||| |||| | |  ::: |  :
Qy   1244 YDRRRDRDEDRDRDRERDRDRDRDRDRDRDRDRDRDRERSRERDRRDRYRD-ERDHRASP 1302

Db    278 rer 280
          |||
Qy   1303 REK 1305


RESULT   15
ENTRY             S28147     #type complete
TITLE             U1 snRNP 70K protein - Arabidopsis thaliana
ALTERNATE_NAMES   U1 small nuclear ribonuleoprotein 70K chain
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse-ear
                    cress
DATE              20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                    12-May-1995
ACCESSIONS        S28147
REFERENCE         S28147
   #authors       Reddy, A.S.N.; Czernik, A.J.; An, G.; Poovaiah, B.W.
   #journal       Biochim. Biophys. Acta (1992) 1171:88-92
   #title         Cloning of the cDNA for U1 small nuclear ribonucleoprotein
                    particle 70K protein from Arabidopsis thaliana.
   #accession     S28147
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-303 ##label RED
CLASSIFICATION    #superfamily ribonucleoprotein repeat homology
FEATURE
   33-100             #domain ribonucleoprotein repeat homology #label RRM
SUMMARY           #length 303  #molecular-weight 36430  #checksum 969

  DB  8;  Score     239;  Match 43.6%;  QryMatch 2.5%;  Pred. No. 5.26e-12;
  Matches    34;  Conservative   13;  Mismatches  27;  Indels   4;  Gaps   4;

Db    167 erersrelsh-eqprersrdrpredkhhrdrdqggrdrdrdsrrdrdrtrdrgdrdrrdr 225
          :|   |     :: |:|  || |:  : ||||:  ||||||  ||||| | | :||||||
Qy   1234 DRKSRRSYHYYDRRRDRDEDRDRDRERDRDRDRD-RDRDRDRDRDRDRERSR-ERDRRDR 1291

Db    226 drgrdrtsrdhdrdrsrk 243
          |  :|  |   |:: ::
Qy   1292 YRD-ERDHRASPREKRQR 1308

Search completed: Wed Jan 17 17:14:59 1996
Job time : 169 secs.
```



Release 2.1D John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jan 17 18:08:52 1996; MasPar time 245.38 Seconds
925.010 Million cell updates/sec

Tabular output not generated.

Title: >US-08-319-745-3
Description: (1:4448) from US08319745.seq
Perfect Score: 4448
N.A. Sequence: 1 GGCCAGTGTCGACACGCTGT.....CTTTTTTTTTTTTTTTTTG 4448
Comp: CCGGTCACAGCTGTGCGACA.....GAAAAAAAAAAAAAAAAAC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-geneseq18
1 part1
2 part2
3 part3
4 part4
5 part5
6 part6
7 part7
8 part8
9 part9
10 part10
11 part11

Statistics: Mean 10.433; Variance 6.001; scale 1.738

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	87	2.0	1047	2	Q10572	Human Natriuretic Pep	1.23e-38
	2	85	1.9	1047	2	Q10572	Human Natriuretic Pep	2.68e-37
	3	46	1.0	204	1	N81164	Base substituted E.co	2.25e-12
c	4	45	1.0	204	1	N81164	Base substituted E.co	8.84e-12

c	5	43	1.0	91	9	Q51746	Oligonucleotide probe	1.33e-10
	6	43	1.0	91	9	Q51746	Oligonucleotide probe	1.33e-10
	7	37	0.8	756	7	Q43704	Sequence encoding a n	3.59e-07
	8	33	0.7	4752	10	Q58032	Tyrosinase promoter.	5.52e-05
	9	33	0.7	3461	10	Q58005	Sequence of plasmid p	5.52e-05
c	10	30	0.7	3699	3	Q23445	Vector pGPe.	2.07e-03
	11	30	0.7	1577	5	Q31726	Rat IgH 3'-enhancer.	2.07e-03
c	12	29	0.7	3871	2	N71302	HSV-1 gB and surround	6.68e-03
c	13	29	0.7	3699	7	Q44179	Plasmid pGPe for clon	6.68e-03
	14	27	0.6	700	9	Q57506	Rat GAP-43 promoter.	6.58e-02
	15	27	0.6	3586	10	Q56643	Human tyrosinase gene	6.58e-02
	16	27	0.6	700	1	Q05263	Sequence of neuronal	6.58e-02
	17	27	0.6	700	9	Q57505	Rat GAP-43 promoter.	6.58e-02
c	18	26	0.6	501	3	N50025	Sequence encoding new	2.00e-01
	19	26	0.6	565	6	Q35072	HCV envelope region n	2.00e-01
	20	25	0.6	501	3	N50025	Sequence encoding new	5.95e-01
c	21	25	0.6	2189	5	Q33127	GABA-A receptor alpha	5.95e-01
c	22	25	0.6	501	3	N50024	Sequence encoding new	5.95e-01
	23	25	0.6	501	3	N50030	Sequence encoding new	5.95e-01
	24	25	0.6	1675	3	Q14634	Clone associated with	5.95e-01
c	25	24	0.5	2312	3	Q22541	KDI-348 DNA.	1.72e+00
c	26	24	0.5	2306	3	Q22544	KDI-354 DNA.	1.72e+00
c	27	24	0.5	2280	3	Q22552	KDI-554 cDNA.	1.72e+00
c	28	24	0.5	501	3	N50023	Sequence encoding new	1.72e+00
c	29	24	0.5	412	2	N60460	Human Factor-VIII-R C	1.72e+00
c	30	24	0.5	501	3	N50028	Sequence encoding new	1.72e+00
c	31	24	0.5	1453	3	Q22548	KDI-I03 cDNA.	1.72e+00
	32	24	0.5	39	7	Q51787	Mixed oligonucleotide	1.72e+00
c	33	24	0.5	1949	3	Q22547	KDI-I320 cDNA.	1.72e+00
c	34	24	0.5	1982	3	Q22546	KDI-I31 cDNA.	1.72e+00
c	35	24	0.5	501	3	N50032	Sequence encoding new	1.72e+00
c	36	24	0.5	1453	3	Q22549	KDI-I07 cDNA.	1.72e+00
	37	24	0.5	501	3	N50024	Sequence encoding new	1.72e+00
c	38	24	0.5	501	3	N50031	Sequence encoding new	1.72e+00
	39	24	0.5	501	3	N50026	Sequence encoding new	1.72e+00
c	40	24	0.5	2319	3	Q22539	KDI-341 DNA.	1.72e+00
c	41	24	0.5	2310	3	Q22542	KDI-350 DNA.	1.72e+00
c	42	24	0.5	2318	3	Q22540	KDI-342 DNA.	1.72e+00
c	43	24	0.5	501	3	N50029	Sequence encoding new	1.72e+00
c	44	24	0.5	2218	3	Q22551	KDI-514 cDNA.	1.72e+00
c	45	24	0.5	2307	3	Q22543	KDI-353 DNA.	1.72e+00

ALIGNMENTS

RESULT 1
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPBR
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047

```
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site  24..26
FT   /label= N-glycos_site
FT   Modified -site  35..37
FT   /label= N-glycos_site
FT   Modified -site  161..163
FT   /label= N-glycos_site
FT   Modified -site  195..197
FT   /label= N-glycos_site
FT   Modified -site  244..246
FT   /label= N-glycos_site
FT   Modified -site  277..279
FT   /label= N-glycos_site
FT   Modified -site  349..351
FT   /label= N-glycos_site
FT   Modified -site  600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC   114,952).  The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence  1047 BP;  87 A;  15 C;  83 G;  51 T;

  DB  2;  Score     87;  Match 8.8%;  QryMatch 2.0%;  Pred. No. 1.23e-38;
  Matches   90;  Conservative  272;  Mismatches 651;  Indels   5;  Gaps   5;

Db    29 vvnnnhnnsyawawnrvgnavanavnangrannvdnrnvssnnngacsnynannsavdnk 88
Cp  3120 CAACAATAGCACCATAACCGCAATGAAGACAGCTGCCAAAGCACACGCCAAAGCCAGTAG 3061

Db    89 nyhdndnnngngcvynaasvarnashwrnnnntagavasgnsaknd hyrtnvrtgnsank 148
Cp  3060 TAAAGATGTCCTTAAATACAAATACTGTTCCCAGAAAAGGAATGGTATACCCGATGGAAA 3001

Db   149 ngnnvvtnhghnnwtaraannyndartddrnhyntnngvnnanngsnnsvnhnvyarnng 208
Cp  3000 GTTCGGTAAACCTTTCGCCTCGTACTTCAGACATAAATCTCGCACAGATCTTATCAACGT 2941

Db   209 gnnnathnnrangrnvyncgnnnmnhnnnnnanrnnntngdyvnnyndvngnsnragn-t 267
Cp  2940 TTTGATGCTATAGTGTCGCTGAGACCGGAAAGGTAGAATGGTAACTGTGTGTAAATTAAT 2881

Db   268 ratgrnwndnrtrnnananrnanntvnvntyrnnnnnnynnnnnrnnnrarndngvnngn 327
Cp  2880 GGCGACGATTTCTTTATTTCTAAATGTACATCCTCCGGAGAATGGATCCATCTTTGGGGC 2821

Db   328 snmnnnagcnydgnnnyanvnnntnnnggtrndgnrnvnkmngrryhgvtgnvvmdknnd 387
```



```
Cp  2820 TGAGGTTTCAAGTTCCCTTGTGAGGCTCCGTATGCCAACGCGTCGTTAGTAGCCCAAGCT 2761

Db   388 rntdnvnwamgdndsgdnnnaahysganknnwwtgrnnnwvkgannsdnnncandnddns 447
Cp  2760 GATAGGTAATTATAAAATGCCTTTGGATTTATAATACCGTCTTTGTCAACTAGTCTGTGA 2701

Db   448 cdktnnstnanvangtgntnnmngvssnnnnrknmnnknnasmnwrnrwnnnnngnsnry 507
Cp  2700 CCTGCCGTAATCAGAGACTTATCGATTGGATTGTCCACATGGCCAGTCTGCACCATAAGT 2641

Db   508 hkgagsrntnsnrgssygsnmtahgkynnnantghnkgnvvankhvnkkrnnntrnvnnn 567
Cp  2640 TTATAGGCCAATATTCCTTCGTCACTCGCGTTTTTGCACCAATACTCTTGTGTTATACAA 2581

Db   568 nkhmrdvnnnhntrnngacndnnnncnvtnycnrgsnndnnnndsnnndwmnrysnnndn 627
Cp  2580 CCGCTGGCAACCTCCTTATCAAAAGCCACTTGCAAGTCCAATAACCAGTCGCGGAATAAA 2521

Db   628 vkgmannhnsnnsshgsnkssncvvdsrnvnkntdygnasnrstannddnnanyakknnt 687
Cp  2520 CTCAACCAAAATTTCGTGAGACCGCCGTTATCATTCTTGATTATATTAGGTATTCTGACG 2461

Db   688 annnnsgnnnnttgmnaadvysngnnnnnnnanrsgnnynngndnsnknnvnkvrngnrny 747
Cp  2460 AATTGATCGTGATACTCATATAATAACTTCTGATTGGTGGGATATTCAAAGTTGCCTTGC 2401

Db   748 nrnsndrtnnnnnnvnnmnrcwandnanrndngnnkgnnrrnnknggtsnndnnnnrmnn 807
Cp  2400 GTCACGGCGTACATATTATAGAAGCCAAAGTATTTTTCCTGACGAGATAAAAATTCGTGT 2341

Db   808 yannnnknvnnrtnaynnnkr-kanannynnnnhsvannnkrgntvnanandsvtnynsd 866
Cp  2340 TCGTCTGTATTCTCCGGTACAATATCAGTCAAATCCAATCCATC-CTTTACTTTTGTCGC 2282

Db   867 nvgntansanstnmnvvtnnndnytcndanndnndvykvntngdaymvvsgnngrngnrh 926
Cp  2281 TCCCCAAACGCTAGTCAGAATAACAGCAATCAACGCTAACATGGATGTAACCTTAACAGC 2222

Db   927 annnarman-anndavssnrnrhrnhdnnrnrngvhtgnvcagvvgnkmnrycnngdtvn 985
Cp  2221 GGGGCGCATGATGAACGGCGCGTATTGGTTCTTGGCCCACTTGGTGAGCGAGACGCTTAA 2162

Db   986 t-asrmnsngnanknhvssttkdandnngcnnnnnrgdvnmkgkgkmrtywnngnrkg 1042
Cp  2161 GCAGCAAGTTTTGGTCACGTTCTCGGACACATCTGGGTCTAGAGGTTGTCTCGTGGTG 2104
```

```
RESULT   2
ID   Q10572 standard; DNA; 1047 BP.
AC   Q10572;
DT   09-APR-1991  (first entry)
DE   Human Natriuretic Peptide Receptor B.
KW   NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW   hyperaldosteronism; glaucoma; guanyl cyclase.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= signal sequence
FT   Protein         12
FT   /label= mature NPBR
FT   Domain          23..455
FT   /label= extracellular domain
FT   /note= "binds natriuretic peptides A,B and C]"
```

```
FT   Domain          456..456
FT   /label= transmembrane domain
FT   Domain          479..1047
FT   /label= cytoplasmic domain
FT   /note= "GC and protien kinase activity"
FT   Modified -site  24..26
FT   /label= N-glycos_site
FT   Modified -site  35..37
FT   /label= N-glycos_site
FT   Modified -site  161..163
FT   /label= N-glycos_site
FT   Modified -site  195..197
FT   /label= N-glycos_site
FT   Modified -site  244..246
FT   /label= N-glycos_site
FT   Modified -site  277..279
FT   /label= N-glycos_site
FT   Modified -site  349..351
FT   /label= N-glycos_site
FT   Modified -site  600..602
FT   /label= N-glycos_site
PN   WO9100292-A.
PD   10-JAN-1991.
PF   22-JUN-1990; U03586.
PR   23-JUN-1989; US-370673.
PA   (GETH ) GENENTECH INC.
PI   Chang M, Goeddel D, Lowe D;
DR   WPI; 91-036711/05.
DR   N-PSDB; Q10324.
PT   Natriuretic protein receptor B - for diagnosis and treatment of
PT   kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS   Claim 3; Fig 1; 49pp; English.
CC   The sequence was derived from the DNA encoding natriuretic peptide
CC   receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC   kinase activity.  The DNA can be inserted into expression vectors
CC   for the prodn. of the protein, opt. after being mutated to produce
CC   NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC   114,952).  The protein (or variants) can be used in treatment of
CC   natriuretic peptide disorders, and also to isolate peptides using
CC   affinity chromatography.  Antibodies with affinity for NPRB can
CC   also be prepd.
SQ   Sequence   1047 BP;    87 A;    15 C;    83 G;    51 T;

  DB  2;  Score      85;  Match 9.0%;  QryMatch 1.9%;  Pred. No. 2.68e-37;
  Matches    83;  Conservative  257;  Mismatches 575;  Indels   9;  Gaps   9;

Db    107 svarnashwrnnnntagavasgnsakndhyrtnvrtgnsankngnnvvtnhghnnwtara 166
Qy   1501 GAAGAACATTATATTAGGCTATATGTTTATGTTAATTTATGTTGCCGTTACTTTAATACA 1560

Db    167 ann-yndartddrnhyntnngvnnanngsnns-vnhnvyarnnggnnnathnnrangrnv 224
Qy   1561 ATGGCGGGATCCGATTCGCTCGCAAGCGGGTGTGGGTATCGCCGGAGTTCTACTACTATC 1620

Db    225 yncgnnnmnhnnnnnanrnnntngdyvnnyndvngnsnragntratgrnwndnrtrnnan 284
Qy   1621 AATCACTGTTGCCGCTGGCTTAGGATTTTGTGCTTTATTAGGCATACCATTCAACGCTTC 1680

Db    285 anrnanntvnvntyrnnnnnnynnnnnrnnnrarndngvnngnsnmnnnagcnydgnnny 344
Qy   1681 AAGTACGCAAATAGTACCATTCCTAGCGCTCGGGTTAGGAGTTCAAGATATGTTTCTTCT 1740

Db    345 anvnnntnnnggtrndgnrnvnkmngrryhgvtgnvvmdknndrntdnvnwamgdndsgd 404
```



```
Qy   1741 CACTCACACGTATGTGGAGCAAGCGGGAGATGTGCCTAGAGAAGAGAGGACTGGACTTGT 1800

Db    405 nnn-aahysganknnwwtgrnnnwvkgannsdnnncandnddnscdktnnstnanvangt 463
Qy   1801 ATTGAAAAAGAGCGGTTTAAGCGTACTTCTGGCGTCTTTGTGCAACGTGATGGCATTTTT 1860

Db    464 gntnnmngvssnnnnrknmnnknnasmnwrnrwnnnnngnsnryhkgagsrntnsnrgss 523
Qy   1861 GGCAGCAGCCCTTCTACCTATTCCAGCTTTCAGAGTATTT-TGCCTACAGGCTGCCATAC 1919

Db    524 ygsnmtahgkynnnantghnkgnvvankhvnkkrnnntrnvnnnnkhmrdvnnnhntrnn 583
Qy   1920 TTCT-TCTGTTTAACTTGGGGTCAATATTACTGGTTTTTCCTGCTATGATCTCGTTAGAC 1978

Db    584 gacndnnnncnvtny-cnrgsnndnnnndsnnndwmnrysnnndnvkgmannhnsnnssh 642
Qy   1979 CTGCGACGGAGGTCAGCCGCGAGGGCCGATCTTTTATGCTGTTTGATGCCTGAGAGTCCA 2038

Db    643 gsnkssncvvdsrnvnkntdygnasnrstannddnnanyakknntannnnsgnnnnttgm 702
Qy   2039 TTACCGAAGAAGAAAATTCCGGAAAGAGCAAAAACTAGAAAAAACGATAAGACTCATAGG 2098

Db    703 naadvysngnnnnnnanrsgnnynngndn-snknnvnkvrngnrnynrnsndrtnnnnnn 761
Qy   2099 ATAGACACCACGAGACAACCTCTAGACCCAGATGTGTCCGAGAACGTGACCAAAACTTGC 2158

Db    762 vnnmnrcwandnanrndngnnkgnnrrnnknggtsnndnnnnrmnnyannnnknvnnrtn 821
Qy   2159 TGCTTAAGCGTCTCGCTCACCAAGTGGGCCAAGAACCAATACGCGCCGTTCATCATGCGC 2218

Db    822 aynnnkrkanannynnnnhsvannnkrgntvnanandsvtnynsdnvgntansanstnmn 881
Qy   2219 CCCGCTGTTAAGGTTACATCCATGTTAGC-GTTGATTGCTGTTATTCTGACTAGCGTTTG 2277

Db    882 vvtnnndnytcndanndnndvykvntngdaymvvsgnngrngnrhannnarmananndav 941
Qy   2278 GGGAGCGACAAAAGTAAAGGATGGATTGGATTTGACTGATATTGTACCGGAGAATACAGA 2337

Db    942 ssnrnrhrnhdnnrnrngvhtgnvcagvvgnkmnrycnngdtvntasrmnsngnanknhv 1001
Qy   2338 CGAACACGAATTTTTAT-CTCGTCAGGAAAAATACTTTGGCTTCTATAATATGTACGCCG 2396

Db   1002 ssttkdandnngcnnnnnrgdvnm 1025
Qy   2397 TGACGCAAGGCAACTTTGAATATC 2420


RESULT    3
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag=  a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
```

```
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  DB  1;  Score      46;  Match 13.6%;  QryMatch 1.0%;  Pred. No. 2.25e-12;
  Matches    14;  Conservative   56;  Mismatches  32;  Indels   1;  Gaps   1;

Db     89 ttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnh 148
          |||::: :::: :  :: :::||| :||:   :::: |:  |:: :::::::::  :: :
Qy    435 TTTCTTCAAGGCGACGCGGGGAAAGTCCTCTTCGTTGCCATCCTCGTTCTGTCGACGTTC 494

Db    149 nncncccbnnhvchn-vhbnnhrnwayvrhdarrddvhccvch 190
            |  | :  :: :  :::  :: :  :::: ::::::| :|:
Qy    495 TGCGTCGGTCTCAAGTCAGCACAAATACATACAAGGGTCGACC 537


RESULT    4
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT   /*tag=  a
FT   /function=multiple cloning site
FT   primer_bind     187..204
FT   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
```



```
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  DB  1;  Score      45;  Match 13.3%;  QryMatch 1.0%;  Pred. No. 8.84e-12;
  Matches    15;  Conservative   60;  Mismatches  37;  Indels   1;  Gaps   1;

Db     73 aayycdchvgccgymrtttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachh 132
          ||::|: :: | |: :   ::  :: : ::::: :::| |:: |::::: ::|||:   ::
Cp   1647 AATCCTAAGCCAGCGGCAACAGTGATTGATAGT-AGTAGAACTCCGGCGATACCCACACC 1589

Db    133 ddhyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvh 185
            :::  :: :: :: :    |||:  :    :::  :: :|::::::| :::::
Cp   1588 CGCTTGCGAGCGAATCGGATCCCGCCATTGTATTAAAGTAACGGCAACATAAA 1536


RESULT    5
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in -
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;   5 A;   17 C;   15 G;   4 T;

  DB  9;  Score      43;  Match 10.9%;  QryMatch 1.0%;  Pred. No. 1.33e-10;
  Matches     6;  Conservative   43;  Mismatches   6;  Indels   0;  Gaps   0;

Db     12 svhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctcaag 66
          ::::  :::::::::::: :::::::::::::::::: ::: ::::: :|||||
Cp   3485 GAACAGCAACCCATCGATGAGTCCCAGAAACACGATGTCCAGTAGTAACCTCAAG 3431


RESULT    6
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
```

```
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  DB  9;  Score      43;  Match 2.0%;  QryMatch 1.0%;  Pred. No. 1.33e-10;
  Matches     1;  Conservative   46;  Mismatches   4;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvc 61
          ::: :::::: :::::::::::::::::::: :::::::: ::::::::::|
Qy    227 CGAGGCTCCTTCGAATCCTCGGATAACGGCTGCACACGAGAGCCCCTGCGC 277


RESULT    7
ID   Q43704 standard; DNA; 756 BP.
AC   Q43704;
DT   29-SEP-1993  (first entry)
DE   Sequence encoding a new cytokine which inhibites induction by gamma
DE   interferon of expression of Class II histocompatibility antigens.
KW   Cytokine; interferon-gamma antagonist; autoimmune disease therapy;
KW   transplants; Class II histocompatibility antigens;
KW   gamma interferon; leukaemia line K562; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             40..528
FT   /*tag=  a
PN   WO9311232-A.
PD   10-JUN-1993.
PF   02-DEC-1992; F01123.
PR   02-DEC-1991; FR-014908.
PA   (INRM ) INSERM INST NAT SANTE & RECH MED.
PI   Augery-bourget Y,  Azzarone B,  Boucheix C,  Jasmin C;
PI   Krief PH;
DR   WPI; 93-197051/24.
DR   P-PSDB; R37991.
PT   New cytokine as interferon-gamma antagonist - inhibits induction
PT   of class II histocompatability antigens on cell surface by
PT   IFN-gamma, for treating and preventing auto:immune disorders
PS   Claim 4; Fig 1; 44pp; French.
CC   The new cytokine specifically inhibits induction by gamma interferon
CC   of expression of Class II histocompatibility antigens on cell
CC   surfaces. It is naturally secreted by cells which are difficult or
CC   impossible to induce with gamme-interferon, e.g. the leukaemia line
CC   K562, from which it can be isolated. To isolate DNA, a K562 gene
CC   bank was constructed in lambda gt11 phage, used to transform Y1090
CC   bacteria and these screened for reactivity with MAbs specific for
CC   the cytokine.Positive colonies were purified,their inserts isolated
CC   with EcoRI and subcloned into pBLSCR KS. Recombinant plasmids were
CC   used to transform HB101 cells and the insert from one positive clone
CC   sequenced. This insert was recloned in vector CDM8 and recombinants
CC   used to transform COS cells. The transformants expressed functional
CC   recombinant cytokine.
SQ   Sequence   756 BP;    254 A;    136 C;    241 G;    125 T;
```



```
  DB  7;  Score      37;  Match 68.0%;  QryMatch 0.8%;  Pred. No. 3.59e-07;
  Matches    70;  Conservative    0;  Mismatches  33;  Indels   0;  Gaps   0;

Db    116 atcgggaacgggagcgtgatcgggaaagagacagagaccgtgaccgagagcgagagcgag 175
          ||||   | |||| || |||   || || ||  ||||||||||  | ||  |||| || |
Qy   3940 ATCGTCGAAGGGATCGCGATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCG 3999

Db    176 aacgagatcgggaacgagagcgagagcgggaccgagagagaga 218
          |  | ||||||||  | || || || |||||| | || || ||
Qy   4000 ACAGGGATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGA 4042


RESULT    8
ID   Q58032 standard; DNA; 4752 BP.
AC   Q58032;
DT   01-AUG-1994  (first entry)
DE   Tyrosinase promoter.
KW   Tyrosinase; cytokine; tumor therapy; gene therapy; melanoma;
KW   B16; tumor-specific promoter; polymerase chain reaction; PCR;
KW   B16; primer; ds.
OS   Mus sp.
PN   WO9404196-A.
PD   03-MAR-1994.
PF   16-AUG-1993; G01730.
PR   14-AUG-1992; GB-017270.
PR   27-FEB-1993; GB-004024.
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI   Hart IR,  Vile RG;
DR   WPI; 94-082848/10.
PT   DNA construct encoding cytokine for expression in tumour cells -
PT   for treating melanoma, pancreatic, breast, colon, prostate cancer
PS   Disclosure; Page 67-70; 107pp; English.
CC   Cytokine genes are expressed in tumor cells, especially melanoma
CC   cells, under the control of the tumor-specific tyrosinase gene
CC   promoter, for gene therapy.  A DNA fragment from the 5' end of
CC   the tyrosinase gene was generated by PCR from genomic DNA of the
CC   B16 melanoma line using the primers given in Q58041-46.
SQ   Sequence   4752 BP;    1550 A;    854 C;    883 G;    1465 T;

  DB 10;  Score      33;  Match 64.6%;  QryMatch 0.7%;  Pred. No. 5.52e-05;
  Matches    73;  Conservative    0;  Mismatches  40;  Indels   0;  Gaps   0;

Db   1799 gaaggagagagagagagagagagagagagagagagagagagagagagagagagagagaga 1858
          || | ||| || ||  ||||  | || || || |||||  | || || ||  | || ||
Qy   3957 GATGAAGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGG 4016

Db   1859 gagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1911
          ||  | ||  | || || || || || |||||  ||  ||||||  |||| ||
Qy   4017 GATCGTGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG 4069


RESULT    9
ID   Q58005 standard; DNA; 3461 BP.
AC   Q58005;
DT   14-SEP-1994  (first entry)
DE   Sequence of plasmid pTHN1 showing the translation elongation factor
DE   (tef1) promoter and coding sequence.
KW   Promoter; translation elongation factor; pTHN1; ss.
OS   Trichoderma reesei.
FH   Key             Location/Qualifiers
FT   promoter        1..1234
FT   /*tag=  a
FT   CDS             1235..1237
FT   /*tag=  b
```

FT /label= start codon
PN WO9404673-A.
PD 03-MAR-1994.
PF 19-AUG-1993; FI0330.
PR 19-AUG-1992; US-932485.
PA (ALKO-) ALKO OY AB.
PI Ilmen MH, Nakari TH, Nevalainen KMH, Onnela M, Pentitilae ME;
DR WPI; 94-083192/10.
PT Cloning promoters active in a desired environmental condition -
PT used partic. for expression of genes in Trichoderma fungal hosts
PT in glucose-contg. medium
PS Claim 15; Figure 1A; 120pp; English.
CC An assay was undertaken to isolate Trichoderma reesei genes which
CC are strongly expressed on glucose. The cDNAs of clones cDNA33,
CC cDNA1, cDNA10, cDNA12 and cDNA 15 were used as probes to isolate
CC corresp. genes and promoters from a Trichoderma chromosomal lambda-
CC bank. The promoters and either the 5' parts of the chromosomal genes
CC or the whole genes were subcloned into pSP73 vector yielding the
CC plasmids pTHN1, pEA33, pTHN3, pEA10, pEA12 and pEA155, corresp. to
CC the clones cDNA33, cDNA1, cDNA10, cDNA12 and cDNA15, respectively.
CC Sequences were obtd. from the 5' ends of the genes and from the
CC promoters using primers designed from previously obtd. sequences.
CC The sequences of the isolated promoters and genes or parts of them
CC (either obtd.from cDNA or chromosomal DNA) are shown in Q58005 for
CC cDNA33, Q58006 for cDNA1, Q58007 for cDNA10, Q58008 for cDNA12,
CC and Q58009 for cDNA15. Based on sequence similarity to known
CC sequences in a protein data bank the clone cDNA33 could be
CC identified as a translation elongation factor, TEF1-alpha.
SQ Sequence 3461 BP; 850 A; 1044 C; 860 G; 697 T;

DB 10; Score 33; Match 82.4%; QryMatch 0.7%; Pred. No. 5.52e-05;
Matches 42; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Db    608 gatcgcgagagggatcactacagggagcgagaccgggacagggatcgcgat 658
          |||||||| ||||||||   | ||||| || ||||||||||||||| | |||
Qy   3993 GATCGCGACAGGGATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGAT 4043
```

RESULT 10
ID Q23445 standard; DNA; 3699 BP.
AC Q23445;
DT 17-AUG-1992 (first entry)
DE Vector pGPe.
KW Mouse; heavy chain; cloning; ss.
OS Rattus rattus.
PN WO9203918-A.
PD 19-MAR-1992.
PF 28-AUG-1991; U06185.
PR 29-AUG-1990; US-574748.
PR 31-AUG-1990; US-575962.
PA (GENP-) GENPHARM INT INC.
PI Lonberg N, Kay R;
DR WPI; 92-113962/14.
PT Immunoglobulin trans:genes - for prodn. of heterologous
PT non-rearranged and/or rearranged Ig chains
PS Example 14; Page 81; 172pp; English.
CC The vector sequence was obtd. by PCR amplification of the
CC immunoglobulin heavy chain 3' enhancer (Patterson, et al., (1990)
CC Nature, 344: 165-168) from rat liver DNA using the PCR primers
CC oligo-44 and oligo-45 (see Q23446,7). The amplified product
CC was digested by BamHI and SphI and cloned into a pUC derived
CC plasmid contg. a polylinker (pNNO3). The resulting plasmid,
CC pRE3, was digested with BamHI and HindIII and the insert contg.
CC the rat Ig heavy chain 3' enhancer cloned into pGPib. The resulting

CC plasmid, pPGe contains several unique restriction sites into which
CC sequences can be cloned and subsequently excised together with the
CC 3' enhancer by NotI digstion, allowing for cloning of very large
CC DNA sequences.
CC See also Q23419-50, Q22417-30.
SQ Sequence 3699 BP; 882 A; 995 C; 899 G; 923 T;

DB 3; Score 30; Match 63.9%; QryMatch 0.7%; Pred. No. 2.07e-03;
Matches 69; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Db    122 ctgtctctgtctctctctgtctctctgtctctgtgtgtgtctctctctgtctctctctct 181
          |||||||  ||||||  |   ||||| || || |   |||| |  ||    || || ||
Cp   4069 CTGTCTCGTTCTCTCGATCGTTCTCTATCCCTATCCCTGTCCCGGTCACGATCCCTATCC 4010

Db    182 gtctctctgtctctctgtctgtctctgtctctgtctctgtctctctct 229
            || ||||| |  | |||||| || || | |||||  || || |||
Cp   4009 CGATCCCTGTCGCGATCTCTGTCCCTTTCACGGTCTCGATCCCTATCT 3962
```

RESULT 11
ID Q31726 standard; DNA; 1577 BP.
AC Q31726;
DT 04-APR-1993 (first entry)
DE Rat IgH 3'-enhancer.
KW Mouse; cross hybridisation; B cell specific; target; ss.
OS Rattus rattus.
PN WO9221762-A.
PD 10-DEC-1992.
PF 03-JUN-1992; SE0375.
PR 07-JUN-1991; SE-001740.
PA (PETT/) PETTERSSON S.
PI Pettersson S;
DR WPI; 92-433660/52.
PT Mouse immunoglobulin H 3' enhancer - which cross-hybridises with
PT rat 3' enhancer and is used to target tissue-specific gene
PT expression
PS Claim 3; Fig 3B 13; 33pp; English.
CC A mouse liver library in bacteriophage was screened using as a probe
CC an AccI-BglI subfragment from the core of the rat 3'enhancer
CC (nucleotides 432-748). Two overlapping gps. of clones (lambda M2 and
CC lambda M3) were isolated. Phage lambda M3 extends 3' of lambda M2.
CC The sequence of the mouse IgH 3' enhancer was determined and aligned
CC with that of the rat. Hybridisations of subclones of the mouse IgH
CC 3' enhancer to phage lambda M2 showed that the 3' enhancers of mouse
CC and rat were present in the genome in opposite orientations. The
CC enhancers are B cell specific and may be used to target tissue
CC specific expression of genes, so may be useful therapeutically, e.g.
CC in targetting prodn. of proteins and in hybridoma technology. The
CC enhancer may be used to enhance expression of genes in host cells, in
CC vivo or in vitro, partic. in certain lymphoid cell lines and in
CC transgenic animals. It may be used for the prodn. of monoclonal
CC antibodies. See also Q31725.
SQ Sequence 1577 BP; 406 A; 402 C; 456 G; 313 T;

DB 5; Score 30; Match 63.9%; QryMatch 0.7%; Pred. No. 2.07e-03;
Matches 69; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Db    837 agagagagacagagacagagacagagacagacagagagacagagagacagagagagac 896
          ||| || ||  ||||| | || || ||||||  |  | |||||  ||    || || ||
Qy   3962 AGATAGGGATCGAGACCGTGAAAGGGACAGAGATCGCGACAGGGATCGGGATAGGGATCG 4021

Db    897 agagagagacacacacagagacagagagacagagagagacagagacag 944
           ||  | ||||   | || || |||||   |  ||||||  |||||||
Qy   4022 TGACCGGGACAGGGATAGGGATAGAGAACGATCGAGAGAACGAGACAG 4069
```

```
RESULT  12
ID  N71302 standard; DNA; 3871 BP.
AC  N71302;
DT  30-APR-1991  (first entry)
DE  HSV-1 gB and surrounding regions.
KW  Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;
KW  glycoprotein; gB; ss.
OS  Herpes simplex virus type 1 (KOS).
FH  Key             Location/Qualifiers
FT  misc_RNA        1..375
FT  /*tag= a
FT  /note= "5' extra sequences beginning with the XhoI
FT  site"
FT  CAAT_signal     406.410
FT  /*tag= b
FT  /number= 1
FT  CAAT_signal     443..448
FT  /*tag= c
FT  /number= 2
FT  TATA_signal     476..479
FT  /*tag= d
FT  misc_RNA        501..789
FT  /*tag= e
FT  /label= mRNA start sequence
FT  /note= "501 is possible start site"
FT  misc_RNA        504..789
FT  /*tag= f
FT  /label= mRNA start sequence
FT  /note= "504 is possible start site"
FT  misc_RNA        506..789
FT  /*tag= g
FT  /label= mRNA start sequence
FT  /note= "506 is possible start site"
FT  CDS             790..3498
FT  /*tag= h
FT  /label= HSV-1 gB
FT  /note= "includes N-terminal hydrophobic leader and
FT  a membrane-spanning sequence, a C-terminal
FT  ionic sequence, and 9 N-linked
FT  saccharide-addition sites"
FT  3'UTR           3499..3549
FT  /*tag= i
FT  polyA_signal    3518..3525
FT  /*tag= j
FT  polyA_site      3549..3549
FT  /*tag= k
FT  misc_RNA        3549..3997
FT  /*tag= l
FT  /note= "3' nonessential sequences to the BamHI
FT  site"
PN  US4642333-A.
PD  10-FEB-1987.
PF  20-JUN-1984; 622496.
PR  16-SEP-1983; US-532996.
PR  20-JUN-1985; US-622496.
PA  (PERS/) PERSON S.
PI  Person S;
DR  WPI; 87-056354/08.
PT  Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as
PT  recombinant and used for vaccines for herpes simplex virus types 1
PT  and 2.
PS  Example; Table 1; 16pp; English.
```



```
CC  !NOTE!  This sequence has been indexed as represented in the
CC  specification, except that bases 'E' have been replaced by 'N'.
CC  The features have been indexed according to the legend of table 1 on
CC  column 19/20 and the Sequence Summary of column 5 (sic).  Note that
CC  the base numbering of the features does not correspond to the
CC  the sequence numbering below.
CC  For another DNA sequence of HSB-1 gB see N71303 (P71135),
CC  and for HSV-2 gB see N71399 (P71136).
CC  A pure non-glycosylated amino acid (AA) chain comprising a sequence
CC  corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
CC  virus which is antigenic to HSV-1 of HSV-2, which contains no more
CC  than 750 AA residues, and which includes AA residues 135-649
CC  inclusive is claimed. It can be used to produce vaccines for
CC  prophylaxis and treatment of HSV-1 and HSV-2.
SQ  Sequence  3871 BP;  743 A;  1402 C;  544 G;  754 T;

  DB  2;  Score      29;  Match 39.6%;  QryMatch 0.7%;  Pred. No. 6.68e-03;
  Matches   53;  Conservative   31;  Mismatches  49;  Indels   1;  Gaps   1;

Db  2817 bcaccatcaccatcacgbtcatccacbccgacbccaacbccbccatsttcbcbbbcbbbc 2876
         :| | ||| |  || ||:|||    |:|   |:| | |:| : |  : ||:|:::|:::
Cp  4042 TCCCTATCCCTGTCCCGGTCACGATCCCTATCCCGATCCCTGTCGCGATCTCTGTCCCTT 3983

Db  2877 bcbcttcttcbabbbgatbbbcgacctbgggcgcgcggtcb-bcaaggtbbtgatbbcat 2935
         :| |    ||:|:::   :::  | | :|  | |   | |: :||  ||  |||| : |
Cp  3982 TCACGGTCTCGATCCCTATCTTCATCGCGATCCCTTCGACGATCATAGTAATGATAGGAA 3923

Db  2936 cbtbbcggtatcbb 2949
         |:|::|| | ||::
Cp  3922 CGTCGCGATTTCCT 3909


RESULT  13
ID  Q44179 standard; DNA; 3699 BP.
AC  Q44179;
DT  10-NOV-1993  (first entry)
DE  Plasmid pGPe for cloning large Ig transgenes.
KW  Immunoglobulin; rat Ig heavy chain 3' enhancer; cloning vector;
KW  heavy chain minilocus transgene; ss.
OS  Synthetic.
FH  Key             Location/Qualifiers
FT  enhancer        1..665
FT  /*tag= a
FT  /note= "rat Ig heavy chain 3' enhancer inserted
FT  between the BamHI and HindIII sites of
FT  the pGP1b polylinker. The 8 Others in this
FT  region correspond to nucleotides which were
FT  illegible in the printed specification"
PN  WO9312227-A.
PD  24-JUN-1993.
PF  17-DEC-1992; U10983.
PR  17-DEC-1991; US-810279.
PR  18-MAR-1992; US-853408.
PR  23-JUN-1992; US-904068.
PA  (GENP-) GENPHARM INT INC.
PI  Kay RM,  Lonberg N;
DR  WPI; 93-214169/26.
PT  Transgenic non-human animals contg. immunoglobulin heavy chain
PT  trans gene - used to produce useful antibodies by isotype
PT  switching
PS  Example 12; Page 90; 196pp; English.
CC  Plasmid pGP1a was derived from pBR322 by insertion of an EcoRI-StyI
CC  linker. The plasmid contains a NotI restriction site downstream
CC  (relative to the ampicillin resistance gene) of a strong
```

```
CC   transcription termination signal derived from the trpA gene.
CC   Potential toxicity of the plasmid is reduced by being of low copy
CC   number (due to retaining the pBR322 copy number control region) and
CC   by eliminating readthrough transcription from the AmpR gene (due to
CC   the strong terminator). Plasmid pGP1b was derived from pGP1a by
CC   insertion of a short polylinker region, flanked by NotI sites.
CC   Oligonucleotides "oligo-44" and "oligo-45" (Q44177 and Q44178,
CC   respectively) were used to amplify the immunoglobulin heavy chain
CC   3' enhancer from rat liver DNA by the polymerase chain reaction
CC   technique. The amplified product was subcloned and an insert
CC   containing the enhancer was cloned into BamHI/HindIII digested
CC   pGP1b. The resulting plasmid pGPe (i.e. Q44179) contains several
CC   unique restriction sites into which large transgene sequences can be
CC   cloned and subsequently excised together with the 3' enhancer.
SQ   Sequence   3699 BP;    882 A;    992 C;    897 G;    920 T;

  DB  7;  Score      29;  Match 63.0%;  QryMatch 0.7%;  Pred. No. 6.68e-03;
  Matches    68;  Conservative    0;  Mismatches  40;  Indels   0;  Gaps   0;

Db    122 ctgtctctgtctctctctgtctctctgtctctgtgtgngtctctctctgtctctctctct 181
          ||||||| ||||||  |   ||||| || || |    ||| |  ||    || || ||
Cp   4069 CTGTCTCGTTCTCTCGATCGTTCTCTATCCCTATCCCTGTCCCGGTCACGATCCCTATCC 4010

Db    182 gtctctctgtctctctgtctgtctctgtctctgtctctgtctctctct 229
            || ||||| |  | |||||| || || | |||||  || || |||
Cp   4009 CGATCCCTGTCGCGATCTCTGTCCCTTTCACGGTCTCGATCCCTATCT 3962


RESULT   14
ID   Q57506 standard; DNA; 700 BP.
AC   Q57506;
DT   09-JUL-1994  (first entry)
DE   Rat GAP-43 promoter.
KW   GAP-43; internal regulatory protein; IRP; recombinant; neuron;
KW   neuronal growth; promoter; ss.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   promoter        1..600
FT   /*tag=  a
FT   /note= "strong potential for H-DNA conformation"
FT   misc_binding    82..89
FT   /*tag=  b
FT   /note= "consensus Pit-1 binding site"
FT   misc_signal     519
FT   /*tag=  c
FT   /note= "transcriptional start site"
FT   misc_signal     545
FT   /*tag=  d
FT   /note= "transcriptional start site"
FT   misc_signal     546
FT   /*tag=  e
FT   /note= "transcriptional start site"
FT   misc_signal     549
FT   /*tag=  f
FT   /note= "transcriptional start site"
FT   misc_signal     550
FT   /*tag=  g
FT   /note= "transcriptional start site"
FT   CDS             601..700
FT   /*tag=  h
FT   /note= "5' end of GAP-43 gene"
PN   AU9347435-A.
PD   23-DEC-1993.
PF   17-SEP-1993; 047435.
```



```
PR   22-DEC-1989; US-465635.
PR   02-JUL-1990; US-546453.
PA   (FEDE/) FEDEROFF H J.
PA   (FISH/) FISHMAN M C.
PA   (STRI/) STRITTMATTER SM.
PA   (VALE/) VALENZUELA D.
PA   (ZUBE/) ZUBER M X.
PI   Federoff HJ,  Fishman MC,  Strittmatter SM,  Valenzuela D;
PI   Zuber MX;
DR   WPI; 94-049278/07.
PT   Recombinant mammalian GAP-43 protein - used to monitor and
PT   regulate neuronal growth in animals, pref. humans
PS   Disclosure; Fig 14; 156pp; English.
CC   The nucleotide sequence of the rat GAP-43 gene promoter region is
CC   given in Q57506.
SQ   Sequence   700 BP;    230 A;    87 C;    278 G;    105 T;

  DB  9;  Score      27;  Match 65.5%;  QryMatch 0.6%;  Pred. No. 6.58e-02;
  Matches    57;  Conservative    0;  Mismatches  30;  Indels   0;  Gaps   0;

Db    280 gggagagagagagagagagagagagagagagagagagagagagagagagagagagaga 339
          |||| |||||  | || || ||  | || || ||  | ||  | || || || || || |
Qy   3985 GGGACAGAGATCGCGACAGGGATCGGGATAGGGATCGTGACCGGGACAGGGATAGGGATA 4044

Db    340 gagagagagagagagagagaaatgc 366
          ||||  ||  ||||||  |||| | ||
Qy   4045 GAGAACGATCGAGAGAACGAGACAGGC 4071


RESULT   15
ID   Q56643 standard; DNA; 3586 BP.
AC   Q56643;
DT   16-SEP-1994  (first entry)
DE   Human tyrosinase gene.
KW   Tyrosinase; albino; albinism; detection; diagnosis;
KW   treatment; prevention; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             987..1805
FT   /*tag=  a
FT   /product= Tyrosinase.
FT   /note= "CDS excludes termination codon."
PN   J06000100-A.
PD   11-JAN-1994.
PF   18-JUN-1992; 185885.
PR   18-JUN-1992; JP-185885.
PA   (DAIL ) DAICEL CHEM IND LTD.
DR   WPI; 94-077141/10.
DR   P-PSDB; R48368.
PT   A DNA sequence which hybridises to the human tyrosinase gene -
PT   used for the detection, diagnosis, prevention and treatment of
PT   human albinism
PS   Disclosure; Page 5-7; 8pp; Japanese.
CC   A DNA sequence having at least 12 successive bases and which
CC   hybridises upstream and/or downstream of the 310th base from the
CC   transcription start site of the human tyrosinase gene can be
CC   used for the detection, diagnosis, prevention and treatment of human
CC   albinism.
SQ   Sequence   3586 BP;    1127 A;    636 C;    760 G;    1063 T;

  DB 10;  Score      27;  Match 65.2%;  QryMatch 0.6%;  Pred. No. 6.58e-02;
  Matches    58;  Conservative    0;  Mismatches  31;  Indels   0;  Gaps   0;

Db    177 gagagagggagagggagagagagggagagagagggagagagagagagagagggagaga 236
```

```
          || || |  ||||   | || |||||  | ||  | ||  | ||  | || ||||| ||
Qy   3981 GAAAGGGACAGAGATCGCCGACAGGGATCGGGATAGGGATCGTGACCGGGACAGGGATAGG 4040

Db    237 gagagagaaagagagagagagggagagag 265
          || |||||| ||  ||||||  |||| ||
Qy   4041 GATAGAGAACGATCGAGAGAACGAGACAG 4069

Search completed: Wed Jan 17 18:13:03 1996
Job time : 251 secs.
```

MA J-X; CHAO J; CHAO L
DEP. BIOCHEMISTRY, MOLECULAR BIOLOGY, MED. UNIV. S.C., 171 ASHLEY AVENUE, CHARLESTON, S.C. 29425, USA.
BIOCHEMISTRY 31 (44). 1992. 10922-10928. CODEN: BICHA
Full Journal Title: Biochemistry
Language: ENGLISH

We have cloned and determined the nucleotide sequence of a novel kallikrein-like mRNA, designated rKlK10*, from rat submandibular gland and kidney with the aid of the polymerase chain reaction (PCR). This cDNA contains 737 base pairs comprising the sequence encoding a mature protein of 235 amino acid residues, partial zymogen peptide, and 3' noncoding sequence. Sequence comparisons showed that rKlk10 mRNA shares 87 and 88% sequence identity with rat tissue kallikrein at nucleic acid and amino acid levels, respectively. It encodes a 26 428-Da acidic protein whose derived amino acid sequence matches completely with the partial amino acid sequence of a kallikrein-like enzyme designated as T-kininogenase, K10 protein, or antigen-.gamma. purified from rat submandibular gland [Xiong et al. (1990) J. Biol. Chem. 265, 2822-2827; Gutman et al. (1991) Eur. J. Biochem. 784, 1-5; Berg et al. (1991) Biochem. J. 280, 19-25]. The protein encoded by rKlk10 retains the key amino acid residues determining kallikrein cleavage specificity. Northern blot analysis with an rKlk10-specific oligonucleotide probe showed that its mRNA level in the submandibular gland is decreased dramatically by administration of the .beta. agonist isoproterenol. Tissue-specific expression of rKlk10 was anlyzed by Northern blotting and Southern blotting of PCR-amplified cDNA, which showed that rKlk10 is expressed at high levels in the submandibular gland and low levels in the kidney but not in seven other tissues including prostate, liver, heart, adrenal gland, testes, pituitary, and pancreas. rKlk10 cDNAs cloned from the kidney and submandibular gland show sequence identity. Specific expression of rKlk10 in the kidney in addition to the submandibular gland indicates that rKlk10 may be involved in the regulation of renal function. The PCR-based cloning strategy provided an efficient and reliable way to potentially identify all expressed kallikrein-related genes in rat submandibular gland.

10/7/33 (Item 6 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

7883667 BIOSIS Number: 40084667
PCR-BASED CLONING STRATEGY FOR G PROTEIN-COUPLED RECEPTORS
PARMENTIER M; LIBERT F; LEFORT A; PERRET J; GERARD C; MAENHAUT C; VAN SANDE J; MOLLEREAU C; EGGERICKX D; ET AL
I.R.I.B.H.N., ULB CAMPUS ERASME, FREE UNIV. BRUSSELS, BRUSSELS, BELG.
SYMPOSIUM ON GENE REGULATION AND SIGNALLING IN ENDOCRINE SYSTEMS HELD AT THE 20TH ANNUAL MEETING OF THE KEYSTONE SYMPOSIA ON MOLECULAR AND CELLULAR BIOLOGY, STEAMBOAT SPRINGS, COLORADO, USA, JANUARY 19-25, 1991. J CELL BIOCHEM SUPPL 0 (15 PART B). 1991. 262. CODEN: JCBSD
Language: ENGLIS

08/319,745

```
  File 154:MEDLINE(R)  1985-1996/Jan W4
         (c) format only 1996 Knight-Ridder Info
  File  55:BIOSIS PREVIEWS(R)  1985-1995/Jan W1
         (c) 1996 BIOSIS
```

& search cloning by PCR

```
      Set  Items  Description
      ---  -----  -----------
?e au=scott m


Ref   Items  Index-term
E1        1  AU=SCOTT LV
E2       10  AU=SCOTT LW
E3      338 *AU=SCOTT M
E4       30  AU=SCOTT M A
E5        3  AU=SCOTT M B
E6       48  AU=SCOTT M C
E7       43  AU=SCOTT M D
E8        3  AU=SCOTT M D S
E9       79  AU=SCOTT M E
E10       5  AU=SCOTT M F
E11     104  AU=SCOTT M G
E12       3  AU=SCOTT M G B

          Enter P or PAGE for more
?p


Ref   Items  Index-term
E13      28  AU=SCOTT M H
E14       1  AU=SCOTT M I H
E15      56  AU=SCOTT M J
E16       2  AU=SCOTT M J III
E17       2  AU=SCOTT M J JR
E18       5  AU=SCOTT M J L
E19      18  AU=SCOTT M K
E20      83  AU=SCOTT M L
E21       7  AU=SCOTT M M
E22       2  AU=SCOTT M N
E23      15  AU=SCOTT M O
E24      96  AU=SCOTT M P

          Enter P or PAGE for more
?p


Ref   Items  Index-term
E25       4  AU=SCOTT M R
E26      10  AU=SCOTT M R D
E27      11  AU=SCOTT M S
E28      11  AU=SCOTT M T
E29       1  AU=SCOTT M-O
E30      16  AU=SCOTT MA
E31       1  AU=SCOTT MB
E32      34  AU=SCOTT MC
E33      27  AU=SCOTT MD
E34      41  AU=SCOTT ME
```

```
E35       64  AU=SCOTT MG
E36       24  AU=SCOTT MH

          Enter P or PAGE for more
?p


Ref    Items  Index-term
E37        2  AU=SCOTT MI
E38        3  AU=SCOTT MILLAR RN
E39       31  AU=SCOTT MJ
E40        8  AU=SCOTT MJ JR
E41        4  AU=SCOTT MJ 3D
E42        3  AU=SCOTT MJ 3RD
E43        6  AU=SCOTT MK
E44       44  AU=SCOTT ML
E45        4  AU=SCOTT MM
E46       10  AU=SCOTT MO
E47        1  AU=SCOTT MOTOR NEURON DIS RES GROUP
E48       63  AU=SCOTT MP

          Enter P or PAGE for more
?s e3 or e24 or e48

              338  AU=SCOTT M
               96  AU=SCOTT M P
               63  AU=SCOTT MP
      S1      497  AU="SCOTT M" OR AU="SCOTT M P" OR AU="SCOTT MP"
?s s1 and patched

              497  S1
              388  PATCHED
      S2        6  S1 AND PATCHED
?rd

...completed examining records
      S3        4  RD (unique items)
?t s3/6/1-4


 3/6/1     (Item 1 from file: 154)
08984026   94299026
  patched  overexpression  causes loss of wingless expression in Drosophila
embryos.


 3/6/2     (Item 2 from file: 154)
07151658   90058658
  The  Drosophila patched gene encodes a putative membrane protein required
for segmental patterning.


 3/6/3     (Item 1 from file: 55)
7236094     BIOSIS Number: 38016615
  GENES THAT CONTROL PATTERN FORMATION DURING DEVELOPMENT


 3/6/4     (Item 2 from file: 55)
```

6834732 BIOSIS Number: 37029111
THE DROSOPHILA SEGMENT POLARITY GENE PATCHED ENCODES A MEMBRANE PROTEIN
?s patched

S4 388 PATCHED
?s s4 and (human or mouse or mosquito or butterfly or beetle)

388 S4
5018157 HUMAN
318853 MOUSE
8288 MOSQUITO
2310 BUTTERFLY
7440 BEETLE
S5 157 S4 AND (HUMAN OR MOUSE OR MOSQUITO OR BUTTERFLY OR BEETLE)
?s s5 and (gene? or clone? or DNA?)

Processing
157 S5
1429785 GENE?
156776 CLONE?
538652 DNA?
S6 14 S5 AND (GENE? OR CLONE? OR DNA?)
?rd

...completed examining records
S7 11 RD (unique items)
?t s7/6/1-11

7/6/1 (Item 1 from file: 154)
09365985 95295985
Human recombinant IGF-I induces the functional expression of AMPA/kainate receptors in cerebellar granule cells.

7/6/2 (Item 2 from file: 154)
09183708 95113708
Surface distribution and partition during freeze-fracture of CD8 antigens on human lymphocytes and on epithelial transfected cells.

7/6/3 (Item 3 from file: 154)
07784014 91303014
Expression and capping of a proliferation-associated surface membrane p34 kDa antigen on different human hematopoietic cell lines.

7/6/4 (Item 4 from file: 154)
07656880 91175880
Lateral diffusion of nerve growth factor receptor: modulation by ligand-binding and cell-associated factors.

7/6/5 (Item 5 from file: 154)
07151658 90058658
The Drosophila patched gene encodes a putative membrane protein required for segmental patterning.

7/6/6 (Item 6 from file: 154)
06796384 89098384
Inhibition of human immunodeficiency virus (HIV-1) replication by synthetic oligo-RNA derivatives.

7/6/7 (Item 7 from file: 154)
05484758 85100758
A kinetic study of membrane immunoglobulin capping by flow cytometry.

7/6/8 (Item 8 from file: 154)
05424908 85040908
Popliteal vein pseudoaneurysm: a case report.

7/6/9 (Item 1 from file: 55)
10525940 BIOSIS Number: 96125940
INTEGRIN-MEDIATED NEURITE OUTGROWTH IN NEUROBLASTOMA CELLS DEPENDS ON THE ACTIVATION OF POTASSIUM CHANNELS

7/6/10 (Item 2 from file: 55)
5778744 BIOSIS Number: 83041051
A METHOD OF MEASURING HUMAN AORTIC VOLUME PULSE WAVE VELOCITY BY ELECTRICAL IMPEDANCE PLETHYSMOGRAPHY AND ITS CLINICAL APPLICATION

7/6/11 (Item 3 from file: 55)
5407358 BIOSIS Number: 82052161
CONSIDERATIONS ON THE TROPHIC NICHES OF TAWNY OWL STRIX-ALUCO AND BARN OWL TYTO-ALBA
?t s7/7/5

7/7/5 (Item 5 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07151658 90058658
The Drosophila patched gene encodes a putative membrane protein required for segmental patterning.
Hooper JE; Scott MP
Howard Hughes Medical Institute, University of Colorado, Boulder 80309-0347.
Cell (UNITED STATES) Nov 17 1989, 59 (4) p751-65, ISSN 0092-8674 Journal Code: CQ4
Contract/Grant No.: HD24584; F32HD06784
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The patched (ptc) gene is one of several segment polarity genes required for correct patterning within every segment of Drosophila. The absence of ptc gene function causes a transformation of the fate of cells in the middle part of each segment so that they form pattern elements characteristic of cells positioned around the segment border. Analysis of the mutant phenotype demonstrates that both segment and parasegment borders

are included in the duplicated pattern of ptc mutants. We have cloned the ptc gene and deduced that the product is a 1286 amino acid protein with at least seven putative transmembrane alpha helices. ptc RNA is expressed in embryos in broad stripes of segmental periodicity that later split into two stripes per segment primordium. The pattern of expression does not directly predict the transformation seen in ptc mutant embryos, suggesting that ptc participates in cell interactions that establish pattern within the segment.

```
?s drosphila and (gal or galactosidase)

            133  DROSPHILA
           7514  GAL
          16555  GALACTOSIDASE
     S8      4  DROSPHILA AND (GAL OR GALACTOSIDASE)
?s drosophila and (gal or galactosidase)

          35618  DROSOPHILA
           7514  GAL
          16555  GALACTOSIDASE
     S9    502  DROSOPHILA AND (GAL OR GALACTOSIDASE)
?s s9 and embryo?

            502  S9
         163692  EMBRYO?
    S10    210  S9 AND EMBRYO?
?s s10a nd develop?

    S11      0  S10A ND DEVELOP?
?s s10 and develop?

            210  S10
         993699  DEVELOP?
    S12    125  S10 AND DEVELOP?
?ds


Set     Items   Description
S1        497   AU="SCOTT M" OR AU="SCOTT M P" OR AU="SCOTT MP"
S2          6   S1 AND PATCHED
S3          4   RD (unique items)
S4        388   PATCHED
S5        157   S4 AND (HUMAN OR MOUSE OR MOSQUITO OR BUTTERFLY OR BEETLE)
S6         14   S5 AND (GENE? OR CLONE? OR DNA?)
S7         11   RD (unique items)
S8          4   DROSPHILA AND (GAL OR GALACTOSIDASE)
S9        502   DROSOPHILA AND (GAL OR GALACTOSIDASE)
S10       210   S9 AND EMBRYO?
S11         0   S10A ND DEVELOP?
S12       125   S10 AND DEVELOP?
?s drosophila(10n)(gal or galactosidase)

          35618  DROSOPHILA
           7514  GAL
          16555  GALACTOSIDASE
    S13    181  DROSOPHILA(10N)(GAL OR GALACTOSIDASE)
?s s13 and develop? and embryo?

            181  S13
```

```
          993699  DEVELOP?
          163692  EMBRYO?
     S14      40  S13 AND DEVELOP? AND EMBRYO?
?rd

...completed examining records
     S15      36  RD (unique items)
?t s15/5/1-36
```

15/5/1 (Item 1 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09502347 96023947

Identification of fat-cell enhancer activity in Drosophila melanogaster using P-element enhancer traps.

Hoshizaki DK; Lunz R; Ghosh M; Johnson W
University of Illinois College of Medicine at Chicago, Department of Biochemistry 60612, USA.

Genome (CANADA) Jun 1995, 38 (3) p497-506, ISSN 0831-2796
Journal Code: FNP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9601
Subfile: INDEX MEDICUS

To identify genes important in fat-cell metabolism and development, we have screened Drosophila stocks carrying an engineered transposable element that can reveal the presence of nearby enhancer elements. We have identified those "enhancer-trap lines" that contain transposable P elements integrated near fat-cell specific enhancer elements. We anticipate that the genes associated with these enhancers will provide information concerning fat-cell function and serve as target genes for studying fat-cell specific gene expression. Furthermore, the identification of enhancer-trap lines active in the developing fat cell should provide an entry point into the molecular and genetic analysis of early fat-cell development. Analysis of two lines has revealed that the transcription factors svp, a steroid-hormone receptor, and Kr, a zinc-finger protein, are present in the fat body; these factors are likely to be involved in fat-cell gene expression. In two other lines, beta-galactosidase was detected in a subset of adepithelial cells that may be the precursors to the adult fat cell. And finally, in a single line transgene activity is present in the progenitor cells of the embryonic fat body. The genes associated with these enhancer-trap lines may be involved in fat-cell development.

Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *Drosophila melanogaster--Physiology--PH; *Enhancer Elements (Genetics)--Genetics--GE; *Fat Body--Physiology--PH; beta-Galactosidase --Metabolism--ME; Drosophila melanogaster--Embryology--EM; Drosophila melanogaster--Genetics--GE; DNA-Binding Proteins; Fat Body--Cytology--CY; Fat Body--Embryology--EM; Genes, Insect; Larva--Growth and Development--GD ; Larva--Genetics--GE; Mesoderm--Cytology--CY; Mesoderm--Physiology--PH; Receptors, Steroid; Transcription Factors--Genetics--GE; Zinc Fingers
CAS Registry No.: 0 (seven-up protein); 0 (DNA-Binding Proteins); 0 (Receptors, Steroid); 0 (Transcription Factors)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: Kr; svp

15/5/2 (Item 2 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09348962 95278962
An improved method for chemical devitellinization of X-gal stained Drosophila embryos.
Singh A; Kango M; Sinha P
Drosophila Stock Center, School of Life Science, Indore, India.
Indian J Exp Biol (INDIA) Feb 1995, 33 (2) p150-2, ISSN 0019-5189
Journal Code: GIZ
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9509
Subfile: INDEX MEDICUS
In Drosophila developmental biological studies, X-gal staining is commonly employed to study the spatio-temporal expression of the lacZ reporter gene in the transformed flies or their embryos. Study of the lacZ pattern in embryos often suffers from the lack of an efficient and high yielding technique for devitellinization of X-gal stained embryos. Devitellinization techniques employed during antibody staining, in situ hybridization or embryonic cuticular preparations generally do not give satisfactory results when used for similar purpose in X-gal stained embryos. This results in the flaky appearance of the blue stain. We present here an improved chemical devitellinization technique which gives a high yield of devitellinized embryos and a better resolution of the X-gal staining pattern.
Tags: Animal; Comparative Study; Support, Non-U.S. Gov't
Descriptors: *Drosophila--Drug Effects--DE; *Galactosides; *Indoles; *Vitelline Membrane--Drug Effects--DE; Drosophila--Embryology--EM; Embryo, Non-Mammalian--Drug Effects--DE; Stains and Staining
CAS Registry No.: 0 (Galactosides); 0 (Indoles); 7240-90-6 (5-bromo-4-chloro-3-indolyl beta-galactoside)

15/5/3 (Item 3 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09337699 95267699
ovo, a Drosophila gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning.
Mevel-Ninio M; Terracol R; Salles C; Vincent A; Payre F
Centre de Genetique Moleculaire du C.N.R.S, Gif sur Yvette, France.
Mech Dev (IRELAND) Jan 1995, 49 (1-2) p83-95, ISSN 0925-4773
Journal Code: AXF
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9508
Subfile: INDEX MEDICUS
Genetic analyses of Drosophila oogenesis have revealed the central role of ovo, a gene required for differentiation of the female germline. A number of recessive ovo mutations also affect the shavenbaby (svb) function required for late embryo patterning, suggesting a tight structural link between ovo and svb. By using various genomic probes for in situ hybridization to wild type and mutant embryos, we show that ovo indeed shares most of its coding sequences with svb. svb expression is detected

early in the presumptive head region and later in each segment. It requires control elements located upstream of the ovo genomic region. ovo expresses abundant maternal RNAs which are uniformly distributed in early cleavage embryos. A fraction that lacks an alternative ovo-specific protein coding region (ORF 2b) is detected in pole cells. Expression of an ovo-specific lacZ reporter gene (ovoB) shows that ovo encodes a nuclear protein present in the germline of both sexes. Zygotic ovoB expression is first detected in embryos at around stage 17 and persists up to the adult stage. Our data show that the germline specific expression of ovo in females correlates with its function in oogenesis. This expression, however, is also observed in males in which ovo is not required.

Tags: Animal; Comparative Study; Female; Male; Support, Non-U.S. Gov't

Descriptors: *Drosophila--Embryology--EM; *Gene Expression Regulation, Developmental--Physiology--PH; *Ovary--Embryology--EM; beta-Galactosidase --Genetics--GE; Amino Acid Sequence; Base Sequence; Drosophila--Genetics --GE; DNA, Recombinant; Embryo, Non-Mammalian; Exons; Genetic Code; Molecular Sequence Data; Ovum--Metabolism--ME; Peptide Chain Initiation; Sequence Homology, Nucleic Acid; Transcription, Genetic

CAS Registry No.: 0 (DNA, Recombinant)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

Gene Symbol: ovo; svb; ovoB; lacZ

15/5/4 (Item 4 from file: 154)

DIALOG(R)File 154:MEDLINE(R)

09241451 95171451

Protein kinase A and hedgehog signaling in Drosophila limb development.

Jiang J; Struhl G

Howard Hughes Medical Institute, Department of Genetics and Development Columbia University College of Physicians and Surgeons, New York, New York 10032.

Cell (UNITED STATES) Feb 24 1995, 80 (4) p563-72, ISSN 0092-8674 Journal Code: CQ4

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9506

Subfile: INDEX MEDICUS

The Drosophila hedgehog (hh) gene encodes a secreted protein involved in organizing growth and patterning in many developmental processes. Hh appears to act by inducing the localized expression of at least two other signaling molecules, decapentaplegic (dpp) and wingless (wg), which then govern cell proliferation and patterning in surrounding tissue. Here, we demonstrate that cyclic AMP (cAMP)-dependent protein kinase A (PKA) is essential during limb development to prevent inappropriate dpp and wg expression. We also show that a constitutively active form of PKA can prevent inappropriate dpp and wg expression, but does not interfere with their normal induction by hh. We propose that the basal activity of PKA imposes a block on the transcription of dpp and wg and that hh exerts its organizing influence by alleviating this block.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Cyclic AMP-Dependent Protein Kinases--Metabolism--ME; *Drosophila--Embryology--EM; *Gene Expression Regulation; *Proteins --Metabolism--ME; *Signal Transduction; beta-Galactosidase--Analysis--AN; Animals, Transgenic; Biological Markers; Cell Communication; Cyclic AMP-Dependent Protein Kinases--Genetics--GE; Down-Regulation (Physiology); Embryonic Induction; Extremities--Embryology--EM; Heat; Insect Hormones

--Biosynthesis--BI; Insect Hormones--Genetics--GE; Membrane Proteins --Genetics--GE; Proteins--Genetics--GE; Proto-Oncogene Proteins --Biosynthesis--BI; Stress; Transcription, Genetic; Wing--Embryology--EM
CAS Registry No.: 0 (dpp protein, Drosophila); 0 (patched protein); 0 (Biological Markers); 0 (Insect Hormones); 0 (Membrane Proteins); 0 (Proteins); 0 (Proto-Oncogene Proteins); 117758-26-6 (wingless protein, Drosophila); 149291-21-4 (hedgehog protein, Drosophila)
Enzyme No.: EC 2.7.10.- (Cyclic AMP-Dependent Protein Kinases); EC 3.2.1.23 (beta-Galactosidase)

15/5/5 (Item 5 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09235092 95165092
Expression of the ligand-binding nicotinic acetylcholine receptor subunit D alpha 2 in the Drosophila central nervous system.
Jonas PE; Phannavong B; Schuster R; Schroder C; Gundelfinger ED
Center for Molecular Neurobiology, University of Hamburg, Germany.
J Neurobiol (UNITED STATES) Dec 1994, 25 (12) p1494-508, ISSN 0022-3034 Journal Code: JAM
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9505
Subfile: INDEX MEDICUS
The D alpha 2 gene encodes a ligand-binding subunit of nicotinic acetylcholine receptors (nAChRs) from Drosophila melanogaster. We have studied the distribution of D alpha 2 transcripts and protein by in situ hybridization and immunohistochemistry, respectively, as well as the regulation of D alpha 2 gene expression in vivo using D alpha 2 promoter fragments fused to the Escherichia coli lacZ gene. Transcripts and protein from the D alpha 2 gene were detected exclusively in the central nervous system. Both in late embryos and adults D alpha 2-like immunoreactivity is widely but not uniformly distributed in the synaptic neuropil, suggesting that the D alpha 2 protein is a subunit of a synaptic nicotinic receptor. Its distribution resembles that of ALS and ARD proteins, two other nAChR subunits of the fly. Five different D alpha 2-lacZ fusion gene constructs were introduced into the Drosophila genome by P-element-mediated gene transfer to identity functional elements of the D alpha 2 promoter. All constructs produce a basic lacZ expression pattern that is compatible with the distribution of D alpha 2 transcripts and protein. A 880 bp upstream fragment harbors the cis elements for the expression of a weak but specific basic D alpha 2 pattern. The next 350 bp further upstream significantly enhance beta-galactosidase expression without influencing the pattern of expression. Between 1.7 and 7.3 kb upstream of the transcription start site one or more elements that are required for D alpha 2 expression in optic lobe tangential cells are located.
Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *beta-Galactosidase--Analysis--AN; *Brain Chemistry; *Drosophila--Genetics--GE; *Gene Expression Regulation, Developmental; *Promoter Regions (Genetics)--Genetics--GE; *Receptors, Nicotinic --Ultrastructure--UL; Base Sequence; Central Nervous System--Cytology--CY; Drosophila--Chemistry--CH; Drosophila--Embryology--EM; In Situ Hybridization; Introns; Molecular Sequence Data; Receptors, Nicotinic --Analysis--AN; Receptors, Nicotinic--Chemistry--CH; Receptors, Nicotinic --Genetics--GE
CAS Registry No.: 0 (Receptors, Nicotinic)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/6 (Item 6 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09233585 95163585
Ultrabithorax protein is necessary but not sufficient for full activation of decapentaplegic expression in the visceral mesoderm.
Sun B; Hursh DA; Jackson D; Beachy PA
Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205.
EMBO J (ENGLAND) Feb 1 1995, 14 (3) p520-35, ISSN 0261-4189
Journal Code: EMB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9505
Subfile: INDEX MEDICUS
To elucidate the mechanisms by which homeotic selector (HOM) genes specify the unique features of Drosophila segments, we have analyzed the regulation of decapentaplegic (dpp), a transforming growth factor (TGF)-beta superfamily member, and have found that the Ultrabithorax (Ubx) HOM protein directly activates dpp expression in parasegment 7 (PS7) of the embryonic visceral mesoderm. Other factors are also required, including one that appears to act through homeodomain protein binding sites and may be encoded by extradenticle (exd). The exd protein binds in a highly co-operative manner to regulatory sequences mediating PS7-specific dpp expression, consistent with a genetic requirement for exd function in normal visceral mesoderm expression of dpp. A second mechanism contributing to PS7 expression of dpp appears not to require Ubx protein directly, and involves a general visceral mesoderm enhancer coupled to a spatially specific repression element. Thus, even in an apparently simple case where visceral mesoderm expression of the dpp target gene mirrors that of the Ubx HOM protein, full activation by Ubx protein requires at least one additional factor. In addition, a distinct regulatory mode not directly involving Ubx protein also appears to contribute to PS7-specific expression.
Tags: Animal
Descriptors: *Drosophila--Embryology--EM; *DNA-Binding Proteins --Metabolism--ME; *Gene Expression Regulation, Developmental; *Homeodomain Proteins--Metabolism--ME; *Insect Hormones--Biosynthesis--BI; beta-Galactosidase--Genetics--GE; Base Sequence; Binding Sites; Carrier Proteins; DNA --Metabolism--ME; Genes, Insect; Genes, Reporter; Insect Hormones--Genetics --GE; Mesoderm--Metabolism--ME; Models, Genetic; Molecular Sequence Data; Regulatory Sequences, Nucleic Acid--Genetics--GE; Sequence Analysis, DNA; Transcription Factors--Metabolism--ME; Transforming Growth Factor beta --Biosynthesis--BI; Viscera--Embryology--EM; Viscera--Metabolism--ME
CAS Registry No.: 0 (dpp protein, Drosophila); 0 (extradenticle protein); 0 (ultrabithorax protein); 0 (Carrier Proteins); 0 (DNA-Binding Proteins); 0 (Homeodomain Proteins); 0 (Insect Hormones); 0 (Transcription Factors); 0 (Transforming Growth Factor beta); 9007-49-2 (DNA)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: ubx; dpp

15/5/7 (Item 7 from file: 154)

DIALOG(R)File 154:MEDLINE(R)

08971560 94286560
Tau-beta-galactosidase, an axon-targeted fusion protein.
Callahan CA; Thomas JB
Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, San Diego, CA 92186.
Proc Natl Acad Sci U S A (UNITED STATES) Jun 21 1994, 91 (13) p5972-6, ISSN 0027-8424 Journal Code: PV3
Contract/Grant No.: GM07198, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9409
Subfile: INDEX MEDICUS
The most commonly used enzymatic reporter molecule, Escherichia coli beta-galactosidase (beta-gal; beta-D-galactoside galactohydrolase, EC 3.2.1.23), fails to readily diffuse into axons; consequently, the morphologies of beta-gal-labeled neurons cannot directly be determined. For analysis of neuronal pathfinding and synaptic connectivity, this information is essential. We have constructed an axon-targeted beta-gal reporter by fusing the cDNA encoding the bovine microtubule-binding protein, tau, to lacZ, the E. coli gene encoding beta-gal. This reporter labels cell bodies and axons when expressed by developing and adult Drosophila neurons. It also reveals the entire cellular extent of nonneuronal cells such as muscle fibers and glia. To generate neuronal markers for studies of Drosophila neural development, we constructed a tau-beta-gal enhancer-trap transposon. From 1500 independent lines generated by mobilization of this transposon, we have isolated a set of useful markers for specific subsets of neurons, glia, and muscles. Since the tau cDNA-lacZ reporter utilizes bovine tau, it may also effectively target beta-gal in vertebrate neurons and prove to be a useful reagent for the analysis of vertebrate nervous systems.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *beta-Galactosidase--Biosynthesis--BI; *tau Proteins --Biosynthesis--BI; *Axons--Physiology--PH; *Neurons--Metabolism--ME; *Recombinant Fusion Proteins--Biosynthesis--BI; beta-Galactosidase --Analysis--AN; tau Proteins--Analysis--AN; Cattle; Drosophila--Embryology --EM; Drosophila--Physiology--PH; Embryo, Non-Mammalian--Physiology--PH; Escherichia coli--Enzymology--EN; Immunohistochemistry; Neurons--Physiology --PH; Organ Specificity; Recombinant Fusion Proteins--Analysis--AN; Restriction Mapping
CAS Registry No.: 0 (tau Proteins); 0 (Recombinant Fusion Proteins)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/8 (Item 8 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08872694 94187694
Multiple cis-acting targeting sequences are required for orb mRNA localization during Drosophila oogenesis.
Lantz V; Schedl P
Department of Biology, Washington University, St. Louis, Missouri 63130.
Mol Cell Biol (UNITED STATES) Apr 1994, 14 (4) p2235-42, ISSN 0270-7306 Journal Code: NGY
Languages: ENGLISH

Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9406
Subfile: INDEX MEDICUS
The targeting of positional information to specific regions of the oocyte or early embryo is one of the key processes in establishing anterior-posterior and dorsal-ventral polarity. In many developmental systems, this is accomplished by localization of mRNAs. The germ line-specific Drosophila orb gene plays a critical role in defining both axes of the developing oocyte, and its mRNA is localized in a complex pattern during oogenesis. We have identified a 280-bp sequence from the orb 3' untranslated region capable of reproducing this complex localization pattern. Furthermore, we have found that multiple cis-acting elements appear to be required for proper targeting of orb mRNA.
Tags: Animal; Female; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila melanogaster--Physiology--PH; *Gene Expression; *Oocytes--Physiology--PH; *Oogenesis; *RNA, Messenger--Biosynthesis--BI; beta-Galactosidase--Biosynthesis--BI; Drosophila melanogaster--Genetics--GE ; Embryo, Non-Mammalian--Physiology--PH; Heat-Shock Proteins--Biosynthesis --BI; Ovary--Physiology--PH; Plasmids; Promoter Regions (Genetics); Recombinant Proteins--Biosynthesis--BI; Restriction Mapping; RNA, Messenger --Analysis--AN; Translation, Genetic
CAS Registry No.: 0 (Heat-Shock Proteins); 0 (Plasmids); 0 (Recombinant Proteins); 0 (RNA, Messenger)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: orb

15/5/9 (Item 9 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08717745 94032745
A double staining technique using 5-bromo,4-chloro-3-indolyl-beta-D-galac topyranoside (X-gal) and immunoperoxidase in whole Drosophila embryos.
Kobayashi S; Okada M
Institute of Biological Sciences, University of Tsukuba, Ibaraki, Japan.
Biotech Histochem (UNITED STATES) Jul 1993, 68 (4) p237-9, ISSN 1052-0295 Journal Code: A29
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9402
Subfile: INDEX MEDICUS
We have developed a double staining method using 5-bromo,4-chloro-3-indolyl-beta-D-galactopyranoside X-gal and immunoperoxid ase for whole Drosophila embryos. The dechorionated embryos are fixed in heptane saturated with 4% formaldehyde, then in heptane and 50% methanol. Fixed embryos are devitellinized with a tungsten needle and processed for immunoperoxidase staining immediately prior to peroxidase color development. The embryos are stained with X-gal, then peroxidase staining is resumed. This procedure enables us to observe cells stained with both X-gal and a specific antibody in whole embryos.
Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *Embryo, Non-Mammalian--Ultrastructure--UL; Drosophila; Galactosidases--Immunology--IM; Galactosides; Immunoenzyme Techniques; Indoles; Paraffin
CAS Registry No.: 0 (Galactosides); 0 (Indoles); 7240-90-6 (5-bromo-4-chloro-3-indolyl beta-galactoside); 8002-74-2 (Paraffin)
Enzyme No.: EC 3.2.1.- (Galactosidases)

15/5/10 (Item 10 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08434171 93144171
The mouse Enhancer trap locus 1 (Etl-1): a novel mammalian gene related to Drosophila and yeast transcriptional regulator genes.
Soininen R; Schoor M; Henseling U; Tepe C; Kisters-Woike B; Rossant J; Gossler A
Max-Delbruck-Laboratorium, Max-Planck-Gesellschaft, Koln, FRG.
Mech Dev (IRELAND) Nov 1992, 39 (1-2) p111-23, ISSN 0925-4773
Journal Code: AXF
Contract/Grant No.: 5R01HD2533403, HD, NICHD
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9305
Subfile: INDEX MEDICUS
A novel mouse gene, Enhancer trap locus 1 (Etl-1), was identified in close proximity to a lacZ enhancer trap integration in the mouse genome showing a specific beta-galactosidase staining pattern during development. In situ analysis revealed a widespread but not ubiquitous expression of Etl-1 throughout development with particularly high levels in the central nervous system and epithelial cells. The amino acid sequence of the Etl-1 protein deduced from the cDNA shows strong similarity, over a stretch of 500 amino acids, to the Drosophila brahma protein involved in the regulation of homeotic genes and to the yeast transcriptional activator protein SNF2/SWI2 as well as to the RAD54 protein and the recently described helicase-related yeast proteins STH1 and MOT1. Etl-1 is the first mammalian member of this group of proteins that are implicated in gene regulation and/or influencing chromatin structure. The homology to the regulatory proteins SNF2/SWI2 and brahma and the expression pattern during embryogenesis suggest that Etl-1 protein might be involved in gene regulating pathways during mouse development.
Tags: Animal; Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *Enhancer Elements (Genetics); *Genes, Structural; *Mice --Genetics--GE; *Proteins--Genetics--GE; beta-Galactosidase--Biosynthesis --BI; beta-Galactosidase--Genetics--GE; Amino Acid Sequence; Base Sequence ; Drosophila melanogaster--Genetics--GE; DNA--Genetics--GE; Fetal Development--Genetics--GE; Genetic Techniques; Mice--Embryology--EM; Molecular Sequence Data; Open Reading Frames; Organ Specificity; Proteins --Biosynthesis--BI; Recombinant Fusion Proteins--Biosynthesis--BI; Recombinant Fusion Proteins--Genetics--GE; RNA, Messenger--Analysis--AN; Saccharomyces cerevisiae--Genetics--GE; Sequence Alignment; Sequence Homology, Amino Acid; Transcription Factors--Biosynthesis--BI; Transcription Factors--Genetics--GE
CAS Registry No.: 0 (Etl-1 protein); 0 (Proteins); 0 (Recombinant Fusion Proteins); 0 (RNA, Messenger); 0 (Transcription Factors); 9007-49-2 (DNA)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: ETL-1; LACZ; SNF2; SW12; MOT1; STH1; brm; neo

15/5/11 (Item 11 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08432018 93142018
A wingless-dependent polar coordinate system in Drosophila imaginal discs [see comments]
Couso JP; Bate M; Martinez-Arias A
Department of Zoology, University of Cambridge, United Kingdom.
Science (UNITED STATES) Jan 22 1993, 259 (5094) p484-9, ISSN 0036-8075 Journal Code: UJ7
Comment in Science 1993 Jan 22;259(5094):471-2
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9304
Subfile: INDEX MEDICUS
The patterning of the imaginal discs in Drosophila melanogaster is a progressive process that, like the patterning of the larval epidermis during embryogenesis, requires the activity of segment polarity genes. One segment polarity gene, wingless, encodes a homolog of the mouse proto-oncogene Wnt-1 and plays a prominent role in the patterning of the larval epidermis and the imaginal discs. However, whereas the function of wingless in the embryo is initially associated with a pattern of stripes along the anteroposterior axis that are part of a Cartesian coordinate system, it is shown here that during imaginal development wingless is associated with a pattern of sectors that provide references for a polar coordinate system homologous to that postulated in a well-known model for the regeneration of insect and vertebrate limbs.
Tags: Animal; Comparative Study; Support, Non-U.S. Gov't
Descriptors: *Drosophila melanogaster--Genetics--GE; beta-Galactosidase --Genetics--GE; Drosophila melanogaster--Embryology--EM; Drosophila melanogaster--Growth and Development--GD; Embryo, Non-Mammalian--Cytology --CY; Embryo, Non-Mammalian--Physiology--PH; Gene Expression; Larva; Mice; Phenotype; Protein-Tyrosine Kinase--Genetics--GE; Proto-Oncogene Proteins --Genetics--GE; Proto-Oncogenes; Sequence Homology, Nucleic Acid; Wing
CAS Registry No.: 0 (proto-oncogene protein int-1); 0 (Proto-Oncogene Proteins)
Enzyme No.: EC 2.7.1.112 (Protein-Tyrosine Kinase); EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: lacZ; wg; en; Wnt-1

15/5/12 (Item 12 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08391146 93101146
Apical secretion and association of the Drosophila yellow gene product with developing larval cuticle structures during embryogenesis.
Kornezos A; Chia W
Drosophila Neurobiology Laboratory, National University of Singapore, Kent Ridge Crescent.
Mol Gen Genet (GERMANY) Nov 1992, 235 (2-3) p397-405, ISSN 0026-8925 Journal Code: NGP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9303
Subfile: INDEX MEDICUS
The yellow (y) gene of Drosophila melanogaster is required for the pigmentation of larval and adult cuticle structures. The deduced y protein sequence includes two putative N-linked glycosylation sites and a putative

signal peptide, suggesting that it might be a secreted molecule. Consistent with the characteristics of a secreted protein, our in vitro translation studies using RNA synthesised from the y cDNA demonstrate that the nascent y polypeptide is a preprotein that cotranslationally translocates into the endoplasmic reticulum (ER) membrane and becomes glycosylated. The N-terminal peptide is cleaved from the preprotein between the two alanine residues at positions 21 and 22, to release the final product into the lumen of the ER. Antibodies raised against the y polypeptide detect the protein starting at 13 h post-fertilization in epidermal cells and in the cuticle structures secreted by them that later become pigmented; in addition, yellow protein is detected in the cuticle structures associated with Keilin's organs. The embryonic beta-galactosidase staining pattern of a transgene, bearing a construct in which expression of the lacZ gene is driven by the y promoter, is also described and is similar to that of the y protein. Our results indicate that the y gene product is an apically secreted protein which becomes an immobilised structural component of the pigmented cuticle.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Drosophila melanogaster--Genetics--GE; *Embryo, Non-Mammalian--Physiology--PH; *Insect Hormones--Genetics--GE; beta-Galactosidase--Genetics--GE; beta-Galactosidase--Metabolism--ME; Amino Acid Sequence; Base Sequence; Drosophila melanogaster--Embryology--EM; DNA --Genetics--GE; Endoplasmic Reticulum--Metabolism--ME; Fertilization; Insect Hormones--Analysis--AN; Insect Hormones--Biosynthesis--BI; Larva; Molecular Sequence Data; Oligodeoxyribonucleotides; Polymerase Chain Reaction--Methods--MT; Recombinant Fusion Proteins--Metabolism--ME; Restriction Mapping; Translation, Genetic

CAS Registry No.: 0 (yellow locus protein, Drosophila); 0 (Insect Hormones); 0 (Oligodeoxyribonucleotides); 0 (Recombinant Fusion Proteins); 9007-49-2 (DNA)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/13 (Item 13 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08219775 92357775

dOct2, a Drosophila Oct transcription factor that functions in yeast.

Prakash K; Fang XD; Engelberg D; Behal A; Parker CS

Division of Chemistry, California Institute of Technology, Pasadena 91125.

Proc Natl Acad Sci U S A (UNITED STATES) Aug 1 1992, 89 (15) p7080-4, ISSN 0027-8424 Journal Code: PV3

Contract/Grant No.: GM42671, GM, NIGMS

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9211

Subfile: INDEX MEDICUS

Oct factors are members of the POU family of transcription factors that are shown to play important roles during development in mammals. Here we report the cDNA cloning and expression of a Drosophila Oct transcription factor. Whole mount in situ hybridization experiments revealed that the spatial expression patterns of this gene during embryonic development have not yet been observed for any other gene. In early embryogenesis, its transcripts are transiently expressed as a wide uniform band from 20% to 40% of the egg length, very similar to that of gap genes. This pattern progressively resolves into a series of narrower stripes followed by

expression in 14 stripes. Subsequently, transcripts from this gene are expressed in the central nervous system and the brain. When expressed in the yeast Saccharomyces cerevisiae, this Drosophila factor functions as a strong, octamer-dependent activator of transcription. Our data strongly suggest possible functions for the Oct factor in pattern formation in Drosophila that might transcend the boundaries of genetically defined segmentation genes.

Tags: Animal; Comparative Study; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: *Drosophila--Genetics--GE; *DNA-Binding Proteins--Metabolism --ME; *Saccharomyces cerevisiae--Genetics--GE; *Transcription Factors --Metabolism--ME; beta-Galactosidase--Genetics--GE; beta-Galactosidase --Isolation and Purification--IP; beta-Galactosidase--Metabolism--ME; Amino Acid Sequence; Base Sequence; Cloning, Molecular; Drosophila --Metabolism--ME; DNA--Genetics--GE; Embryo, Non-Mammalian; Gene Library; Molecular Sequence Data; Plasmids; Protein Conformation; Recombinant Fusion Proteins--Isolation and Purification--IP; Recombinant Fusion Proteins --Metabolism--ME; Sequence Homology, Nucleic Acid; Transcription Factors --Genetics--GE; Transcription Factors--Isolation and Purification--IP

Molecular Sequence Databank No.: GENBANK/M93149

CAS Registry No.: 0 (transcription factor OTF-2); 0 (DNA-Binding Proteins); 0 (Plasmids); 0 (Recombinant Fusion Proteins); 0 (Transcription Factors); 9007-49-2 (DNA)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/14 (Item 14 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08209352 92347352

Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo.

Jiang J; Rushlow CA; Zhou Q; Small S; Levine M

Department of Biological Sciences, Fairchild Center, Columbia University, New York, NY 10027.

EMBO J (ENGLAND) Aug 1992, 11 (8) p3147-54, ISSN 0261-4189 Journal Code: EMB

Contract/Grant No.: GM 46638, GM, NIGMS

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9211

Subfile: INDEX MEDICUS

The dorsal (dl) morphogen gradient is responsible for initiating the differentiation of the mesoderm, neuroectoderm and dorsal ectoderm in the Drosophila embryo. dl encodes a sequence-specific DNA binding protein that belongs to the Rel family of transcription factors. Previous studies have shown that dl activates the mesoderm determinant twist (twi); here we use a combination of site-directed mutagenesis and P-transformation assays to demonstrate that it also functions as a direct transcriptional repressor of a second target gene, zerknullt (zen). By exchanging dl binding sites between the promoters we show that activator sites from twi can mediate repression when placed in the context of the zen promoter, and that repressor sites from zen can mediate activation in the context of the twi promoter. This represents the first demonstration that common binding sites for any DNA binding protein can mediate both activation and repression in a developing embryo. Evidence is also presented that the affinities of dl binding sites are important for the efficiency of repression, but are not

the sole determinants of the threshold response to the dl gradient.
Tags: Animal; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *DNA-Binding Proteins--Genetics --GE; *Nuclear Proteins--Genetics--GE; beta-Galactosidase--Genetics--GE; beta-Galactosidase--Metabolism--ME; Base Sequence; Binding Sites; Cloning, Molecular; Drosophila--Embryology--EM; DNA Insertion Elements; DNA-Binding Proteins--Metabolism--ME; Embryo, Non-Mammalian--Physiology--PH; Heat-Shock Proteins--Genetics--GE; Heat-Shock Proteins--Metabolism--ME; Molecular Sequence Data; Morphogenesis--Genetics--GE; Mutagenesis, Site-Directed; Nuclear Proteins--Metabolism--ME; Oligodeoxyribonucleotides; Promoter Regions (Genetics); Recombinant Fusion Proteins--Metabolism--ME; Transformation, Genetic
CAS Registry No.: 0 (dorsal protein, Drosophila); 0 (DNA Insertion Elements); 0 (DNA-Binding Proteins); 0 (Heat-Shock Proteins); 0 (Nuclear Proteins); 0 (Oligodeoxyribonucleotides); 0 (Recombinant Fusion Proteins)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: dl; zen; lacZ

15/5/15 (Item 15 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08090796 92228796
Dorsal-ventral patterning in Drosophila: DNA binding of snail protein to the single-minded gene.
Kasai Y; Nambu JR; Lieberman PM; Crews ST
Department of Biology, University of California, Los Angeles 90024.
Proc Natl Acad Sci U S A (UNITED STATES) Apr 15 1992, 89 (8) p3414-8, ISSN 0027-8424 Journal Code: PV3
Contract/Grant No.: T 32 CA09056, CA, NCI; R01 HD25251, HD, NICHD
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9207
Subfile: INDEX MEDICUS
The Drosophila snail gene is required for proper mesodermal development. Genetic studies suggest that it functions by repressing adjacent ectodermal gene expression including that of the single-minded (sim) gene. The snail gene encodes a protein with a zinc-finger motif, and here we report that the snail gene product is a sequence-specific DNA binding protein. The snail protein recognizes a 14-base-pair consensus sequence that is found nine times in a 2.8-kilobase sim regulatory region. These results provide evidence for the direct control of sim transcription by snail.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *DNA-Binding Proteins--Genetics --GE; *Genes; *Zinc Fingers--Genetics--GE; beta-Galactosidase--Genetics--GE ; beta-Galactosidase--Metabolism--ME; Alleles; Base Sequence; Blastoderm --Physiology--PH; Drosophila--Embryology--EM; DNA-Binding Proteins --Metabolism--ME; Embryo, Non-Mammalian--Physiology--PH; Molecular Sequence Data; Oligodeoxyribonucleotides; Oncogene Proteins, Viral--Genetics--GE; Promoter Regions (Genetics)
CAS Registry No.: 0 (Adenovirus Early Proteins); 0 (DNA-Binding Proteins); 0 (Oligodeoxyribonucleotides); 0 (Oncogene Proteins, Viral)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: sim; Sna

15/5/16 (Item 16 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08073453 92211453
Functional assay of a putative Drosophila sodium channel gene in homozygous deficiency neurons.
Germeraad S; O'Dowd D; Aldrich RW
Department of Biological Sciences, San Jose State University, CA 95192.
J Neurogenet (ENGLAND) Feb 1992, 8 (1) p1-16, ISSN 0167-7063
Journal Code: JKE
Contract/Grant No.: NS27501, NS, NINDS; NS23294, NS, NINDS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9207
Subfile: INDEX MEDICUS
Using voltage-clamp techniques, we have examined embryonic sodium currents in neurons deficient for a gene located at 60E5/6 that shares extensive amino acid similarity with vertebrate sodium channel genes. These neurons expressed sodium currents similar to wildtype, supporting the hypothesis that para, and not the gene at 60E5/6, is the primary sodium channel gene expressed in embryonic neurons. A simple marking procedure allowing positive identification of the genotypes of cultured Drosophila embryos obtained from heterozygous parents was used to recognize cultures homozygous for deficiencies. The morphological development of both neurons and myotubes in these cultures was similar to wildtype, making it feasible to compare the properties of normal diploid cells and cells completely lacking a putative sodium channel gene.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *Neurons--Physiology--PH; *Sodium Channels--Physiology--PH; beta-Galactosidase--Analysis--AN; beta-Galactosidase--Genetics--GE; Cells, Cultured; Chromosome Deletion; Chromosome Mapping; Drosophila--Embryology--EM; Drosophila--Physiology--PH; Homozygote; Ion Channel Gating--Genetics--GE; Larva; Mutation--Genetics--GE; Phenotype; Sequence Homology, Nucleic Acid
CAS Registry No.: 0 (Sodium Channels)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/17 (Item 17 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08026858 92164858
Regulatory interactions and role in cell type specification of the Malpighian tubules by the cut, Kruppel, and caudal genes of Drosophila.
Liu S; Jack J
Program in Molecular Biology, Sloan-Kettering Institute, New York, New York.
Dev Biol (UNITED STATES) Mar 1992, 150 (1) p133-43, ISSN 0012-1606
Journal Code: E7T
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9206
Subfile: INDEX MEDICUS
Kruppel and caudal genes are both required for normal segmentation of the embryo, and the developmental regulatory gene cut is necessary for the normal specification of external sensory organs. These three genes are also

expressed in the Malpighian tubules before and during differentiation. Two of the genes, Kruppel and cut, are known to be required for development of the tubules. We report that the absence of maternal and zygotic caudal function reduces their normal growth and elongation. Normal Kruppel function, which is known to be required for caudal expression, is also required for cut expression, while cut and caudal are expressed independently of each other. Cell type transformations of Malpighian tubules were studied by examining the effects of mutations on the expression of markers specific to Malpighian tubules, hindgut, or midgut of normal embryos. Loss of Kruppel activity confers hindgut characteristics on those cells that normally form the Malpighian tubules with all markers tested. Loss of cut function alters the expression of some markers but not others. The pathway of tissue specific gene regulation, apparently, branches beyond Kruppel to form at least a cut and a caudal branch.

Tags: Animal; Comparative Study; Support, U.S. Gov't, Non-P.H.S.

Descriptors: *Cell Differentiation--Genetics--GE; *Drosophila--Genetics --GE; *Gene Expression Regulation, Enzymologic--Genetics--GE; *Malpighian Tubules--Embryology--EM; beta-Galactosidase--Analysis--AN; Drosophila --Embryology--EM; DNA--Analysis--AN; Malpighian Tubules--Metabolism--ME; Mutation

CAS Registry No.: 9007-49-2 (DNA)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

Gene Symbol: cut; Kruppel; caudal

15/5/18 (Item 18 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08018168 92156168

A single locus encodes both phenylalanine hydroxylase and tryptophan hydroxylase activities in Drosophila.

Neckameyer WS; White K

Department of Biology, Brandeis University, Waltham, Massachusetts 02254.

J Biol Chem (UNITED STATES) Feb 25 1992, 267 (6) p4199-206, ISSN 0021-9258 Journal Code: HIV

Contract/Grant No.: NS23510, NS, NINDS; RRO 4671, RR, NCRR

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9205

Subfile: INDEX MEDICUS

We have used a full-length clone encoding rabbit tryptophan hydroxylase (TRH) to isolate the Drosophila homologue (DTPH). Southern analysis of Drosophila genomic DNA reveals a pattern indicative of a single gene. The single transcript is expressed in adult head and body mRNA but is also detected in mRNA from early embryos. The embryonic transcript is ubiquitously expressed and appears to concentrate in yolk granules. In situ hybridization of TRH-homologous antisense RNA probe to sectioned tissue from third instar larvae demonstrated the presence of this transcript in fat body and cuticular tissue. Developmental immunoblot analysis using antibodies raised against a beta-galactosidase-Drosophila fusion protein revealed a 45-kDa embryonic protein also detected in female abdomens and a 50-kDa protein found in larval and adult stages. Immunocytochemical analysis of the Drosophila protein in the larval central nervous system showed that it appeared to be present in both serotonin- and catecholamine-containing neurons. A nonfusion protein generated in Escherichia coli hydroxylates both tryptophan and phenylalanine. We propose that there are only two aromatic amino acid hydroxylase genes in

Drosophila: one encoding tyrosine hydroxylase, DTH, and DTPH, a gene encoding both tryptophan and phenylalanine hydroxylase activities.
Tags: Animal; Female; Male; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila melanogaster--Genetics--GE; *Phenylalanine Hydroxylase--Genetics--GE; *Tryptophan Hydroxylase--Genetics--GE; beta-Galactosidase--Metabolism--ME; Amino Acid Sequence; Base Sequence; Blotting, Western; Chromosome Mapping; Drosophila melanogaster--Enzymology--EN; DNA--Genetics--GE; Electrophoresis, Polyacrylamide Gel; Gene Expression; Immunohistochemistry; Molecular Sequence Data; Nucleic Acid Hybridization; Phenylalanine Hydroxylase--Metabolism--ME; Recombinant Fusion Proteins--Metabolism--ME; Restriction Mapping; RNA--Genetics--GE; Tryptophan Hydroxylase--Metabolism--ME
Molecular Sequence Databank No.: GENBANK/M81833; GENBANK/X59129; GENBANK/X59130; GENBANK/L01671; GENBANK/L01672; GENBANK/L01673; GENBANK/L01674; GENBANK/L01675; GENBANK/L01676; GENBANK/L01677
CAS Registry No.: 0 (Recombinant Fusion Proteins); 0 (RNA); 9007-49-2 (DNA)
Enzyme No.: EC 1.14.16.1 (Phenylalanine Hydroxylase); EC 1.14.16.4 (Tryptophan Hydroxylase); EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: DTH; DTPH

15/5/19 (Item 19 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07828894 91347894
An extensive 3' cis-regulatory region directs the imaginal disk expression of decapentaplegic, a member of the TGF-beta family in Drosophila.
Blackman RK; Sanicola M; Raftery LA; Gillevet T; Gelbart WM
Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138-2097.
Development (ENGLAND) Mar 1991, 111 (3) p657-66, ISSN 0950-1991 Journal Code: ECW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9112
Subfile: INDEX MEDICUS
The decapentaplegic (dpp) gene in Drosophila melanogaster encodes a TGF-beta-like signalling molecule that is expressed in a complex and changing pattern during development. One of dpp's contributions is to proximal-distal outgrowth of the adult appendages, structures derived from the larval imaginal disks. Appendage specific mutations of dpp fall in a 20 kb interval 3' to the known dpp transcripts. Here, we directly test the hypothesis that these mutations define an extended 3' cis-regulatory region. By analysis of germ-line transformants expressing a reporter gene, we show that sequences from this portion of the gene, termed the dppdisk region, are capable of directing expression comparable to that defined by RNA in situ hybridization. We localize two intervals of the dppdisk region that appear to account for much of the dpp spatial pattern in imaginal disks and discuss the positions of these important elements in terms of the genetics of dpp. Finally, we provide evidence to suggest that one of our constructs expresses beta-galactosidase in the early imaginal disk primordia in the embryo, at approximately the time when they are set aside from surrounding larval epidermal tissues. Thus, dpp may be involved directly in the determination of the imaginal disks.
Tags: Animal; Female; Male; Support, Non-U.S. Gov't; Support, U.S. Gov't,

P.H.S.
Descriptors: *Drosophila--Genetics--GE; *DNA; *Gene Expression--Genetics --GE; *Genes, Structural--Physiology--PH; *Regulatory Sequences, Nucleic Acid--Physiology--PH; beta-Galactosidase--Genetics--GE; Animals, Transgenic ; Drosophila--Embryology--EM; Drosophila--Ultrastructure--UL; Enhancer Elements (Genetics)--Genetics--GE; Microscopy, Electron; Transforming Growth Factor beta--Genetics--GE
CAS Registry No.: 0 (Transforming Growth Factor beta); 9007-49-2 (DNA)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/20 (Item 20 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07811452 91330452
Generating lineage-specific markers to study Drosophila development.
Perrimon N; Noll E; McCall K; Brand A
Department of Genetics, Harvard Medical School, Boston, MA 02115.
Dev Genet (UNITED STATES) 1991, 12 (3) p238-52, ISSN 0192-253X
Journal Code: DEG
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9111
Subfile: INDEX MEDICUS
To generate cell- and tissue-specific expression patterns of the reporter gene lacZ in Drosophila, we have generated and characterized 1,426 independent insertion strains using four different P-element constructs. These four transposons carry a lacZ gene driven either by the weak promoter of the P-element transposase gene or by partial promoters from the even-skipped, fushi-tarazu, or engrailed genes. The tissue-specific patterns of beta-galactosidase expression that we are able to generate depend on the promoter utilized. We describe in detail 13 strains that can be used to follow specific cell lineages and demonstrate their utility in analyzing the phenotypes of developmental mutants. Insertion strains generated with P-elements that carry various sequences upstream of the lacZ gene exhibit an increased variety of expression patterns that can be used to study Drosophila development.
Tags: Animal; Female; Male; Support, Non-U.S. Gov't
Descriptors: *Drosophila--Genetics--GE; *Genetic Markers--Genetics--GE; beta-Galactosidase--Genetics--GE; Cloning, Molecular; Drosophila --Embryology--EM; DNA Insertion Elements; Gene Expression; Immunoenzyme Techniques; Mutation; Organ Specificity--Genetics--GE; Promoter Regions (Genetics)
CAS Registry No.: 0 (DNA Insertion Elements); 0 (Genetic Markers)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/21 (Item 21 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07546324 91065324
Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis.
Simon J; Peifer M; Bender W; O'Connor M
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

EMBO J (ENGLAND) Dec 1990, 9 (12) p3945-56, ISSN 0261-4189
Journal Code: EMB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9103
Subfile: INDEX MEDICUS

The Drosophila bithorax complex (BX-C) controls segmental development by selectively deploying three protein products, Ubx, abd-A and Abd-B, within specific segments along the body axis. Expression of these products within any one segment (or, more accurately, parasegment) is affected by mutations clustered in a particular region of the BX-C. The regulatory regions defined by this genetic analysis span 20-50 kb and there is one region for each segmental unit. Here we describe regulatory elements from several of these regions, identified by fusion to a Ubx-lacZ gene and analysis in germline transformants. A small DNA fragment from the abx region programs expression with an anterior boundary in the second thoracic segment (parasegment 5). This anterior limit is appropriate, since the abx region normally controls Ubx in parasegment 5. Other regulatory regions of the BX-C that control development of parasegments 6, 7 or 8 contain similar regulatory elements that program expression with anterior limits in parasegments 6, 7 or 8, respectively. These experiments define a class of BX-C regulatory elements that control expression along the anterior-posterior axis. The early appearance of the lacZ patterns in embryos suggests a role for these elements in the initial activation of expression from the BX-C.

Tags: Animal; Female; Male; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: *Drosophila--Genetics--GE; *Genes, Regulator; beta-Galactosidase--Genetics--GE; beta-Galactosidase--Metabolism--ME; Crosses, Genetic; Drosophila--Anatomy and Histology--AH; Drosophila--Embryology--EM; Embryo, Non-Mammalian--Physiology--PH; Enhancer Elements (Genetics); Genes, Homeobox; Genetic Vectors; Recombinant Fusion Proteins--Biosynthesis--BI; Thorax--Anatomy and Histology--AH

CAS Registry No.: 0 (Recombinant Fusion Proteins)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)
Gene Symbol: Ubx

15/5/22 (Item 22 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07527017 91046017
Evidence for positive and negative regulation of the Hox-3.1 gene.
Bieberich CJ; Utset MF; Awgulewitsch A; Ruddle FH
Department of Biology, Yale University, New Haven, CT 06511.
Proc Natl Acad Sci U S A (UNITED STATES) Nov 1990, 87 (21) p8462-6, ISSN 0027-8424 Journal Code: PV3
Contract/Grant No.: GM 009966, GM, NIGMS; GM 43334-01, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9102
Subfile: INDEX MEDICUS

The region-specific patterns of expression of mouse homeobox genes are considered important for establishing the embryonic body plan. A 5-kilobase (kb) DNA fragment from the Hox-3.1 locus that is sufficient to confer region-specific expression to a beta-galactosidase reporter gene in transgenic mouse embryos has been defined. The observed reporter gene

expression pattern closely parallels endogenous Hox-3.1 expression in 8- to 9.5-day postcoitum (p.c.) embryos. At 10.5 days p.c. and later, the pattern of beta-galactosidase activity diverges from the Hox-3.1 pattern, and an inappropriately high level of reporter gene expression is observed in posterior spinal ganglia. Inclusion of an additional 2 kb of upstream sequences is sufficient to suppress this aberrant expression in the developing spinal ganglia. Together, these results show that the control of early Hox-3.1 expression is complex, involving at least one positively acting and one negatively acting element.

Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: *Gene Expression Regulation; *Genes, Homeobox; beta-Galactosidase--Genetics--GE; Drosophila--Genetics--GE; Embryo --Enzymology--EN; Escherichia coli--Enzymology--EN; Escherichia coli --Genetics--GE; Mice; Mice, Transgenic; Restriction Mapping

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

Gene Symbol: Hox-3.1; lacZ

15/5/23 (Item 23 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07387149 90294149

P-element-mediated enhancer detection allows rapid identification of developmentally regulated genes and cell specific markers in Drosophila.

Bellen HJ; Wilson C; Gibson G; Grossniklaus U; Pearson RK; O'Kane C; Gehring WJ

Dept. of Cell Biology, Biozentrum, Univ. of Basel, Switzerland.

J Physiol (Paris) (FRANCE) 1990, 84 (1) p33-41, ISSN 0021-7948

Journal Code: JRB

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9010

Subfile: INDEX MEDICUS

We have employed a new technique in Drosophila that allows in vivo detection of genomic regulatory elements using a beta-galactosidase reporter gene. A translational fusion of the reporter gene to the P-transposase gene, which is encoded by the P-transposon of Drosophila, places the expression of beta-galactosidase under the control of the weak P-transposase promoter. Flies carrying single insertions of this P-element construct at different locations in the Drosophila genome frequently stain for beta-galactosidase activity in a temporally and spatially restricted fashion in embryos, larvae and adult ovaries, reflecting the influence of nearby genomic regulatory elements on the P-transposase promoter. This technique is a powerful tool as it can be used to produce very many different cell markers and to isolate developmentally regulated genes in Drosophila. We discuss the implications of our results and the applications of the technique to further the study of Drosophila development.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Biological Markers--Analysis--AN; *Drosophila melanogaster --Genetics--GE; *DNA Insertion Elements; *Enhancer Elements (Genetics); *Regulatory Sequences, Nucleic Acid; beta-Galactosidase--Analysis--AN; Drosophila melanogaster--Embryology--EM; Embryo, Non-Mammalian--Analysis --AN; Gene Expression Regulation; Oogenesis--Genetics--GE

CAS Registry No.: 0 (Biological Markers); 0 (DNA Insertion Elements)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/24 (Item 24 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07201680 90108680
P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila.
Wilson C; Pearson RK; Bellen HJ; O'Kane CJ; Grossniklaus U; Gehring WJ
Department of Cell Biology, Biozentrum, University of Basel, Switzerland.
Genes Dev (UNITED STATES) Sep 1989, 3 (9) p1301-13, ISSN 0890-9369 Journal Code: FN3
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9004
Subfile: INDEX MEDICUS
We describe a new approach for identifying and studying genes involved in Drosophila development. Single copies of an enhancer detector transposon, P[1ArB], have been introduced into flies at many different genomic locations. The beta-galactosidase reporter gene in this construct is influenced by a wide range of genomic transcriptional regulatory elements in its vicinity. Our results suggest that a significant proportion of these regulatory sequences are control elements of nearby Drosophila genes. These genes need not be disrupted for their regulatory elements to be identified by P[1ArB]. The P[1ArB] transposon has been designed to facilitate both rapid cloning and deletion analysis of genomic sequences into which it inserts. Therefore, the enhancer detection system is an efficient method of screening for genes primarily on the basis of their expression pattern and then rapidly analyzing those of particular interest at the molecular and genetic levels.
Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *Drosophila melanogaster--Genetics--GE; *DNA Insertion Elements; *Gene Expression Regulation; *Regulatory Sequences, Nucleic Acid ; beta-Galactosidase--Analysis--AN; Drosophila melanogaster--Embryology--EM ; Embryo, Non-Mammalian--Analysis--AN; Fetal Development; Genetic Vectors; Recombinant Fusion Proteins--Analysis--AN
CAS Registry No.: 0 (DNA Insertion Elements); 0 (Recombinant Fusion Proteins)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/25 (Item 25 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07201679 90108679
P-element-mediated enhancer detection: a versatile method to study development in Drosophila.
Bellen HJ; O'Kane CJ; Wilson C; Grossniklaus U; Pearson RK; Gehring WJ
Department of Cell Biology, Biozentrum, University of Basel, Switzerland.
Genes Dev (UNITED STATES) Sep 1989, 3 (9) p1288-300, ISSN 0890-9369 Journal Code: FN3
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 9004
Subfile: INDEX MEDICUS
We generated and characterized greater than 500 Drosophila strains that carry single copies of a novel P-element enhancer detector. In the majority of the strains, the beta-galactosidase reporter gene in the P-transposon

responds to nearby transcriptional regulatory sequences in the genome. A remarkable diversity of spatially and temporally regulated staining patterns is observed in embryos carrying different insertions. We selected numerous strains as markers for different embryonic organs, tissues, and cells. Many of these strains should allow the study of complex developmental processes, such as nervous system development, which have not been convenient to analyze previously. Also, we present genetic evidence that some of the detected regulatory elements control nearby Drosophila genes. In light of our results, we discuss the diversity and complexity of cis-acting regulatory elements in the genome and the general applications of the enhancer detector method for the study of Drosophila development.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Drosophila melanogaster--Genetics--GE; *DNA Insertion Elements; *Enhancer Elements (Genetics); beta-Galactosidase--Analysis--AN; Biological Markers--Analysis--AN; Drosophila melanogaster--Embryology--EM; Embryo, Non-Mammalian--Analysis--AN; Gene Expression Regulation; Genetic Vectors; Recombinant Fusion Proteins--Analysis--AN

CAS Registry No.: 0 (Biological Markers); 0 (DNA Insertion Elements); 0 (Recombinant Fusion Proteins)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/26 (Item 26 from file: 154)

DIALOG(R)File 154:MEDLINE(R)

07201678 90108678

Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector.

Bier E; Vaessin H; Shepherd S; Lee K; McCall K; Barbel S; Ackerman L; Carretto R; Uemura T; Grell E; et al

Howard Hughes Medical Institute, San Francisco, California.

Genes Dev (UNITED STATES) Sep 1989, 3 (9) p1273-87, ISSN 0890-9369 Journal Code: FN3

Languages: ENGLISH

Document type: JOURNAL ARTICLE

JOURNAL ANNOUNCEMENT: 9004

Subfile: INDEX MEDICUS

A P-element vector has been constructed and used to generate lines of flies with single autosomal P-element insertions. The lines were analyzed in two ways: (1) the identification of cis-acting patterning information within the Drosophila genome, as revealed by a lacZ reporter gene within the P element, and (2) the isolation of lethal mutations. We examined 3768 independent lines for the expression of lacZ in embryos and looked among these lines for lethal mutations affecting embryonic neurogenesis. This type of screen appears to be an effective way to find new loci that may play a role in the development of the Drosophila nervous system.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: *Drosophila melanogaster--Genetics--GE; *DNA Insertion Elements; beta-Galactosidase--Analysis--AN; Drosophila melanogaster --Embryology--EM; Embryo, Non-Mammalian--Analysis--AN; Evolution; Gene Expression Regulation; Genes, Lethal; Genetic Vectors; Mutation; Nervous System--Embryology--EM; Recombinant Fusion Proteins--Analysis--AN

CAS Registry No.: 0 (DNA Insertion Elements); 0 (Recombinant Fusion Proteins)

Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/27 (Item 27 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07082425 89384425
Regulatory elements involved in the tissue-specific expression of the yellow gene of Drosophila.
Martin M; Meng YB; Chia W
Department of Biochemistry, School of Medical Sciences, University of Bristol, UK.
Mol Gen Genet (GERMANY, WEST) Jul 1989, 218 (1) p118-26, ISSN 0026-8925 Journal Code: NGP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8912
Subfile: INDEX MEDICUS
We have assessed the DNA sequence requirements for the correct spatial pattern and phenotypic expression of y in the late embryo/larvae. The wild-type larval phenotype requires both the regions between -294 bp and -92 bp and a portion of the intron; the sequence element(s) located within the intron can act in a position independent manner to effect the wild-type larval phenotype. The larval expression pattern was examined by tissue experiments in situ and by staining germline transformants derived from various y/lacZ fusion constructs. The larval expression of y is restricted to the mouthparts, microsetae and anal plates. While the -495 bp to +194 bp region alone cannot effect a wild-type larval expression pattern, this region in conjunction with the intron appears to be sufficient to drive beta-gal expression in an essentially wild-type pattern. Our data further suggest that the -294 bp to -92 bp region contains elements which specify the larval pattern and that the element(s) in the intron normally act to enhance the level of expression necessary for the wild-type larval phenotype. We also present a phenotypic analysis of the adult cuticle structures of germline transformants derived from a variety of deletion and rearrangement constructs of the y gene. This analysis has revealed several new features associated with the regulation of y expression.
Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *Drosophila--Genetics--GE; *Insect Hormones--Genetics--GE; *Regulatory Sequences, Nucleic Acid; beta-Galactosidase--Metabolism--ME; Base Sequence; Cloning, Molecular; Drosophila--Growth and Development--GD; DNA Insertion Elements; Gene Expression Regulation; Insect Hormones --Biosynthesis--BI; Introns; Lac Operon; Nucleic Acid Hybridization; Phenotype; Promoter Regions (Genetics); Restriction Mapping; Transcription, Genetic; Transformation, Genetic
CAS Registry No.: 0 (yellow locus protein, Drosophila); 0 (DNA Insertion Elements); 0 (Insect Hormones)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/28 (Item 28 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06894899 89196899
Developmental expression of the Drosophila zeste gene and localization of zeste protein on polytene chromosomes.
Pirrotta V; Bickel S; Mariani C
Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030.

Genes Dev (UNITED STATES) Dec 1988, 2 (12B) p1839-50, ISSN 0890-9369 Journal Code: FN3
Contract/Grant No.: GM-34630
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8907
Subfile: INDEX MEDICUS

The expression of the zeste gene varies through the life cycle of the fly. Its transcription is most abundant in maternal RNA, declines to very low levels during larval growth, but rises again in late third instar larvae and pupae. Using transposons containing a zeste-lacZ gene, we found a corresponding variation in the tissue distribution of zeste from stage to stage. Nearly ubiquitous expression of the zeste-lacZ gene is found in late embryos and first instar larvae, but disappears almost completely except in brain and gonads by third instar larva. Shortly before pupation expression rises again in imaginal discs, Malpighian tubules, and salivary glands and again becomes nearly ubiquitous in pupae. zeste continues to be expressed in adult brain and gonads. We constructed flies carrying a zeste gene controlled by the heat shock promoter and studied the distribution of zeste protein in their polytene chromosomes as well as those of wild-type flies. Using affinity-purified anti-zeste antibodies, we find that wild-type salivary gland chromosomes contain about 60 strong bands of zeste immunofluorescence at specific cytological locations. After heat induction of larvae containing the hs-zeste gene, many hundreds of bands appear. These results suggest the involvement of zeste in the expression of a wide variety of genes at different developmental stages.

Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *Gene Expression Regulation; *Genes, Structural; *Insect Hormones--Genetics--GE; beta-Galactosidase --Genetics--GE; Cloning, Molecular; Drosophila--Growth and Development--GD; DNA Insertion Elements; Fluorescent Antibody Technique; Heat-Shock Proteins --Genetics--GE; Larva--Genetics--GE; Mutation; Pupa--Genetics--GE; Salivary Glands--Cytology--CY
CAS Registry No.: 0 (DNA Insertion Elements); 0 (Heat-Shock Proteins) ; 0 (Insect Hormones)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/29 (Item 29 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06809103 89111103
Expression of an engrailed-like gene during development of the early embryonic chick nervous system.
Gardner CA; Darnell DK; Poole SJ; Ordahl CP; Barald KF
Department of Anatomy and Cell Biology, University of Michigan, Ann Arbor 48109.
J Neurosci Res (UNITED STATES) Oct-Dec 1988, 21 (2-4) p426-37, ISSN 0360-4012 Journal Code: KAC
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8905
Subfile: INDEX MEDICUS

The engrailed gene has been identified in Drosophila as an important developmental gene involved in the control of segmentation. Here we describe the embryonic expression of a chicken gene, ChickEn (Darnell et

al.: J Cell Biol 103(5):311a, 1986), which contains homology to the Drosophila engrailed gene. Northern blots of early chick embryo tissue poly(A)+ RNA resulted in hybridization to at least three bands expressed predominantly in the brain/head region when probed with ChickEn genomic fragments. Eight cDNA clones generated from embryonic day 6 (stage 29-30) chick brain poly(A)+ RNA are identical in their nucleotide sequence with the ChickEn genomic clone. In situ hybridization to sections of 4-day (stage 24) embryos indicated that ChickEn transcripts were concentrated in the posterior mesencephalon and anterior metencephalon. In cultures of chick cranial neural crest cells (eight to nine somites; stage 9) ChickEn transcripts were localized in a subset (approx. 8%) of cells examined after 2 days in culture. A mouse monoclonal antibody, inv-4D9D4, made by Coleman and Kornberg recognizes the engrailed-like homeo domain of the engrailed and invected proteins (Martin-Blanco, Coleman, and Kornberg, personal communication). Patel, Coleman, Kornberg and Goodman (unpublished) have shown that this antibody binds to the hindbrain of 2-day-old chick embryos. We have confirmed these results and shown that this antibody binds to the same region of 4-day (stage 24) chick brains that in situ hybridization showed contained ChickEn transcripts. This antibody also recognizes a homeo domain-containing ChickEn peptide expressed as a beta-galactosidase fusion protein in Drosophila cell culture. We have not detected ChickEn protein in any tissue prior to eight to nine somites (stage 9). These results delineate the major expression pattern of the ChickEn gene during early (prior to stage 30) embryonic development in the chick.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Descriptors: *Fetal Development; *Gene Expression Regulation; *Genes, Structural; *Neural Crest--Metabolism--ME; Chick Embryo; Chromosome Mapping ; Nucleic Acid Hybridization; RNA--Metabolism--ME
CAS Registry No.: 0 (RNA)

15/5/30 (Item 30 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06805994 89107994
Control elements of the P2 promoter of the Antennapedia gene.
Boulet AM; Scott MP
Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder 80302.
Genes Dev (UNITED STATES) Dec 1988, 2 (12A) p1600-14, ISSN 0890-9369 Journal Code: FN3
Contract/Grant No.: 18163
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8905
Subfile: INDEX MEDICUS
Antennapedia (Antp), a homeotic gene of Drosophila required for proper differentiation of the thorax of the fly, is expressed in complex spatial patterns during development. The gene is greater than 100 kb long and has two independently regulated promoters. To characterize cis-acting regulatory elements responsible for the expression pattern, fusions of the Antp promoter 2 cap site and upstream sequences to an Adh-lacZ gene were introduced into flies. A 10-kb sequence directs beta-galactosidase production in a pattern that closely resembles the endogenous P2 pattern. Transcription from the 10-kb fusions is regulated by three genes that regulate Antp transcription. Control elements, including a target of action

of homeo-domain-containing proteins, were mapped by deleting parts of the 10-kb sequence.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *Genes, Homeobox; *Promoter Regions (Genetics); *Regulatory Sequences, Nucleic Acid; beta-Galactosidase --Biosynthesis--BI; beta-Galactosidase--Genetics--GE; Cloning, Molecular; Drosophila--Embryology--EM; Escherichia coli--Genetics--GE; Mutation; Plasmids; Recombinant Fusion Proteins--Biosynthesis--BI; Recombinant Fusion Proteins--Genetics--GE; Restriction Mapping; Transformation, Genetic
CAS Registry No.: 0 (Plasmids); 0 (Recombinant Fusion Proteins)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/31 (Item 31 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06551887 88196887
Spatial and temporal expression of the period gene in Drosophila melanogaster.
Liu X; Lorenz L; Yu QN; Hall JC; Rosbash M
Department of Biology, Brandeis University, Waltham, Massachusetts 02254.
Genes Dev (UNITED STATES) Feb 1988, 2 (2) p228-38, ISSN 0890-9369 Journal Code: FN3
Contract/Grant No.: GM33205
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8808
Subfile: INDEX MEDICUS
The temporal and spatial expression of the period gene of Drosophila melanogaster has been analyzed by examining the expression of a per beta-galactosidase fusion gene in transformants and by in situ hybridization experiments with wild-type flies. Several strains of Drosophila melanogaster, transformed with the fusion gene, have been generated. The gene is active in mid-late embryos in the midline of the nervous system. Thereafter, beta-galactosidase activity is undetectable until the pupal stage when the prothoracic gland-corpora allata and the optic lobes are beta-galactosidase positive. In adults a surprisingly large number of tissues stain positively, including antennae, proboscis, eyes, optic lobes, cells of the central brain, cells of the thoracic ganglia, gut, Malpighian tubules, and ovarian follicle cells. The temporal pattern of expression agrees well with previous estimates made from developmental Northern blots with RNA extracted from wild-type animals. We suggest that many of the tissues that express the per gene contain their own intrinsic oscillator activity.
Tags: Animal; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila melanogaster--Genetics--GE; *Gene Expression Regulation; beta-Galactosidase--Genetics--GE; Circadian Rhythm; Cloning, Molecular; Drosophila melanogaster--Growth and Development--GD; Drosophila melanogaster--Physiology--PH; Mutation; Transformation, Genetic
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/32 (Item 32 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06452440 88097440

Detection in situ of genomic regulatory elements in Drosophila.
O'Kane CJ; Gehring WJ
Department of Cell Biology, University of Basel, Switzerland.
Proc Natl Acad Sci U S A (UNITED STATES) Dec 1987, 84 (24) p9123-7, ISSN 0027-8424 Journal Code: PV3
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8804
Subfile: INDEX MEDICUS
We have developed an approach for the in situ detection of genomic elements that regulate transcription zin Drosophila melanogaster. The approach is analogous to a powerful method of bacterial genetics, the random generation of operon fusions, that enables the isolation and characterization of genes simply by knowing or postulating their pattern of expression; it is not necessary initially to screen for mutant phenotypes. To apply this approach to Drosophila, we have used the expression of the lacZ gene of Escherichia coli from the P-element promoter in germ-line transformant flies to screen for chromosomal elements that can act at a distance to stimulate expression from this apparently weak promoter. Of 49 transformed fly lines obtained, approximately 70% show some type of spatially regulated expression of the lacZ gene in embryos; many of these express lacZ specifically in the nervous system. The P-lacZ fusion gene is, therefore, an efficient tool for the recovery of elements that may regulate gene expression in Drosophila and for the generation of a wide variety of cell-type-specific markers.
Tags: Animal; Support, Non-U.S. Gov't
Descriptors: *Drosophila melanogaster--Genetics--GE; *DNA Insertion Elements; *Gene Expression Regulation; *Promoter Regions (Genetics); *Regulatory Sequences, Nucleic Acid; beta-Galactosidase--Diagnostic Use--DU ; Drosophila melanogaster--Embryology--EM; DNA, Recombinant; Nervous System --Physiology--PH; Tissue Distribution
CAS Registry No.: 0 (DNA Insertion Elements); 0 (DNA, Recombinant)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/33 (Item 33 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

06313209 87287209
Borders of parasegments in Drosophila embryos are delimited by the fushi tarazu and even-skipped genes.
Lawrence PA; Johnston P; Macdonald P; Struhl G
Nature (ENGLAND) Jul 30-Aug 5 1987, 328 (6129) p440-2, ISSN 0028-0836 Journal Code: NSC
Languages: ENGLISH
Document type: JOURNAL ARTICLE
JOURNAL ANNOUNCEMENT: 8711
Subfile: INDEX MEDICUS
One of the earliest molecular signs of segmentation in Drosophila embryos is the striped expression of some pair-rule genes during the blastoderm stage. Two of these genes, fushi-tarazu (ftz) and even-skipped (eve) are expressed during this stage in complementary patterns of seven stripes which develop and disappear in concert. Here, we map the cells expressing each of these two pair-rule genes with respect to the 14 stripes of cells expressing the engrailed gene. We find that both ftz and eve generate stripes which have sharp boundaries at the anterior margin, but fade away posteriorly. The anterior boundaries correspond cell by cell with the

anterior boundaries of expression of the engrailed gene. We therefore suggest that a key function of early ftz and eve gene activity is the formation of a sharp stable boundary at the anterior margin of each stripe. These boundary lines, rather than the narrowing zonal stripes, would delimit the anterior boundaries of engrailed and other homoeotic genes and thereby subdivide the embryo into parasegments.
Tags: Animal; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Descriptors: *Drosophila--Genetics--GE; *Genes, Homeobox; beta-Galactosidase--Genetics--GE; Blastoderm--Metabolism--ME; Blastoderm--Ultrastructure--UL; Drosophila--Embryology--EM; DNA, Recombinant; Gene Expression Regulation; Histocytochemistry; Immunologic Tests
CAS Registry No.: 0 (DNA, Recombinant)
Enzyme No.: EC 3.2.1.23 (beta-Galactosidase)

15/5/34 (Item 1 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

11510853 BIOSIS Number: 98110853
Regulation of DNA replication-related gene expression during Drosophila development
Yamaguchi M; Hirose F; Matsukage A
Lab. Cell Biol., Aichi Cancer Cent Res. Inst., Nagoya 464, Japan
Cell Structure and Function 19 (6). 1994. 463.
Full Journal Title: Forty-seventh Annual Meeting of the Japan Society for Cell Biology, Nagasaki, Japan, September 28-30, 1994. Cell Structure and Function
ISSN: 0386-7196
Language: ENGLISH
Document Type: CONFERENCE PAPER
Print Number: Biological Abstracts/RRM Vol. 047 Iss. 003 Ref. 045215
Descriptors/Keywords: MEETING ABSTRACT; DROSOPHILA; PROLIFERATING CELL NUCLEAR ANTIGEN; MESSENGER RNA; BETA-GALACTOSIDASE; PROMOTER; EMBRYOGENESIS; DEVELOPMENT
Concept Codes:
*03506 Genetics and Cytogenetics-Animal
*10300 Replication, Transcription, Translation
*10808 Enzymes-Physiological Studies
*25502 Developmental Biology-Embryology-General and Descriptive
*25508 Developmental Biology-Embryology-Morphogenesis, General
*34502 Immunology and Immunochemistry-General; Methods
*64076 Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology-Insecta-Physiology
00520 General Biology-Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals
10062 Biochemical Studies-Nucleic Acids, Purines and Pyrimidines
10064 Biochemical Studies-Proteins, Peptides and Amino Acids
Biosystematic Codes:
75314 Diptera
Super Taxa:
Animals; Invertebrates; Arthropods; Insects

15/5/35 (Item 2 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

10405441 BIOSIS Number: 96005441
DEVELOPMENTAL REGULATORY ELEMENTS IN THE 5' FLANKING DNA OF THE DROSOPHILA CHOLINE ACETYLTRANSFERASE GENE
KITAMOTO T; SALVATERRA P M
DIV. NEUROSCI., BECKMAN RESEARCH INST. CITY HOPE, 1450 EAST DUARTE RD., DUARTE, CA 91010, USA.
ROUX'S ARCH DEV BIOL 202 (3). 1993. 159-169. CODEN: RADBE
Full Journal Title: Roux'S Archives of Developmental Biology
Language: ENGLISH
Choline acetyltransferase (ChAT, EC 2.3.1.6) catalyzes the production of the neurotransmitter acetylcholine, and is an essential factor for neurons to be cholinergic. We have analyzed regulation of the Drosophila ChAT gene during development by examining the .beta.-galactosidase expression pattern in transformed lines carrying different lengths of 5' flanking DNA fused to a lacZ reporter gene. The largest fragment tested, 7.4 kb, resulted in the most extensive expression pattern in embryonic and larval nervous system and likely reflects all the cis-regulatory elements necessary for ChAT expression. We also found that 5' flanking DNA located between 3.3 kb and 1.2 kb is essential for the receptor gene expression in most of the segmentally arranged embryonic sensory neurons as well as other distinct cells in the CNS. The existence of negative regulatory elements was suggested by the observation that differentiating photoreceptor cells in eye imaginal discs showed the reporter gene expression in several 1.2 kb and 3.3 kb transformants but not in 7.4 kb transformants. Furthermore, we have fused the 5' flanking DNA fragments to a wild type ChAT cDNA and used these constructs to transform Drosophila with a Cha mutant background. Surprisingly, even though different amounts of 5' flanking DNA resulted in different spatial expression patterns, all of the positively expressing cDNA transformed lines were rescued from lethality. Our results suggest that developmental expression of the ChAT gene is regulated both positively and negatively by the combined action of several elements located in the 7.4 kb upstream region, and that the more distal 5' flanking DNA is not necessary for embryonic survival and development to adult flies.
Descriptors/Keywords: EMBRYO LARVAE ACETYLCHOLINE CHOLINE ACETYLTRANSFERASE EC 2.3.1.6 NEURON CENTRAL NERVOUS SYSTEM
Concept Codes:
*02506 Cytology and Cytochemistry-Animal
*03506 Genetics and Cytogenetics-Animal
*10062 Biochemical Studies-Nucleic Acids, Purines and Pyrimidines
*10808 Enzymes-Physiological Studies
*20504 Nervous System-Physiology and Biochemistry
*25502 Developmental Biology-Embryology-General and Descriptive
*25508 Developmental Biology-Embryology-Morphogenesis, General
*64076 Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology-Insecta-Physiology
10064 Biochemical Studies-Proteins, Peptides and Amino Acids
Biosystematic Codes:
75314 Diptera
Super Taxa:
Animals; Invertebrates; Arthropods; Insects

15/5/36 (Item 3 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

10086391 BIOSIS Number: 95086391
A CIS-ELEMENT MEDIATING ULTRABITHORAX AUTOREGULATION IN THE CENTRAL

NERVOUS SYSTEM
CHRISTEN B; BIENZ M
MRC LABORATORY MOLECULAR BIOLOGY, HILLS ROAD, CAMBRIDGE CB2 2QH, UK.
MECH DEV 39 (1-2). 1992. 73-80. CODEN: MEDVE
Language: ENGLISH
We dissected an upstream control region (a BXD fragment) from the homeotic gene Ultrabithorax (Ubx) of Drosophila which confers a Ubx-like expression pattern in the embryonic ectoderm. We found several distinct enhancer elements spread through the whole BXD fragment each of which is active in transformed embryos, mediating a different pattern of .beta.-galactosidase expression in the ventral nerve cord. The strongest of these patterns mimics Ubx expression within the Ubx domain. This pattern is strictly dependent on Ubx function. Thus, the BXD control region contains a Ubx response element, suggesting that positive autoregulation of Ubx may occur in the central nervous system of the developing embryo.
Descriptors/Keywords: DROSOPHILA EMBRYO BETA GALACTOSIDASE VENTRAL NERVE CORD TRANSCRIPTIONAL ACTIVATION DEVELOPMENT
Concept Codes:
*03506 Genetics and Cytogenetics-Animal
*10300 Replication, Transcription, Translation
*10808 Enzymes-Physiological Studies
*20504 Nervous System-Physiology and Biochemistry
*25502 Developmental Biology-Embryology-General and Descriptive
*25508 Developmental Biology-Embryology-Morphogenesis, General
*64076 Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology-Insecta-Physiology
10064 Biochemical Studies-Proteins, Peptides and Amino Acids
Biosystematic Codes:
75314 Diptera
Super Taxa:
Animals; Invertebrates; Arthropods; Insects
?s patched and (gal or galactosidase)

```
            388  PATCHED
           7514  GAL
          16555  GALACTOSIDASE
    S16      4  PATCHED AND (GAL OR GALACTOSIDASE)
```

?t s16/6/1-4

16/6/1 (Item 1 from file: 154)
09241451 95171451
Protein kinase A and hedgehog signaling in Drosophila limb development.

16/6/2 (Item 2 from file: 154)
07664999 91183999
The Drosophila segment polarity gene patched is involved in a position-signalling mechanism in imaginal discs.

16/6/3 (Item 3 from file: 154)
05982167 86283167
The influence of serosal patch size on the growth of small intestinal neomucosa.

16/6/4 (Item 1 from file: 55)

7767662 BIOSIS Number: 90135662
THE DROSOPHILA SEGMENT POLARITY GENE PATCHED IS INVOLVED IN A POSITION-SIGNALING MECHANISM IN IMAGINAL DISCS
?t s16/7/1-3

16/7/1 (Item 1 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09241451 95171451
Protein kinase A and hedgehog signaling in Drosophila limb development.
Jiang J; Struhl G
Howard Hughes Medical Institute, Department of Genetics and Development Columbia University College of Physicians and Surgeons, New York, New York 10032.
Cell (UNITED STATES) Feb 24 1995, 80 (4) p563-72, ISSN 0092-8674 Journal Code: CQ4
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The Drosophila hedgehog (hh) gene encodes a secreted protein involved in organizing growth and patterning in many developmental processes. Hh appears to act by inducing the localized expression of at least two other signaling molecules, decapentaplegic (dpp) and wingless (wg), which then govern cell proliferation and patterning in surrounding tissue. Here, we demonstrate that cyclic AMP (cAMP)-dependent protein kinase A (PKA) is essential during limb development to prevent inappropriate dpp and wg expression. We also show that a constitutively active form of PKA can prevent inappropriate dpp and wg expression, but does not interfere with their normal induction by hh. We propose that the basal activity of PKA imposes a block on the transcription of dpp and wg and that hh exerts its organizing influence by alleviating this block.

16/7/2 (Item 2 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07664999 91183999
The Drosophila segment polarity gene patched is involved in a position-signalling mechanism in imaginal discs.
Phillips RG; Roberts IJ; Ingham PW; Whittle JR
School of Biological Sciences, University of Sussex, Brighton, UK.
Development (ENGLAND) Sep 1990, 110 (1) p105-14, ISSN 0950-1991 Journal Code: ECW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We demonstrate the role of the segment polarity gene patched (ptc) in patterning in the cuticle of the adult fly. Genetic mosaics of a lethal allele of patched show that the contribution of patched varies in a position-specific manner, defining three regions in the wing where ptc clones, respectively, behave as wild-type cells, affect vein formation, or are rarely recovered. Analysis of twin clones demonstrates that the reduced clone frequency results from a proliferation failure or cell loss. In the region where clones upset venation, they autonomously fail to form veins and also non-autonomously induce ectopic veins in adjacent wild-type cells. In heteroallelic combinations with lethal alleles, two viable alleles produce distinct phenotypes: (1) loss of structures and mirror-image

duplications in the region where patched clones fail to proliferate; (2) vein abnormalities in the anterior compartment. We propose that these differences reflect independently mutable functions within the gene. We show the pattern of patched transcription in the developing imaginal wing disc in relation to the expression of certain other reporter genes using a novel double-labelling method combining non-radioactive detection of in situ hybridization with beta-galactosidase detection. The patched transcript is present throughout the anterior compartment, with a stripe of maximal intensity along the A/P compartment border extending into the posterior compartment. We propose that the patched product is a component of a cell-to-cell position-signalling mechanism, a proposal consistent with the predicted structure of the patched protein.

16/7/3 (Item 3 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

05982167 86283167
The influence of serosal patch size on the growth of small intestinal neomucosa.
Bragg LE; Thompson JS
J Surg Res (UNITED STATES) May 1986, 40 (5) p426-31, ISSN 0022-4804
Journal Code: K7B
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Several factors might affect the growth of neomucosa after serosal patching of small intestinal defects. Often only short segments of small intestine can be patched because of limited serosal surface and anatomic factors. The purpose of this study was to determine the influence of patch size on neomucosal growth. Twenty male New Zealand white rabbits underwent patching with colon serosa of either a 2 X 15-cm distal ileal defect (n = 10) or three 2 X 5-cm ileal defects (n = 10). There was significantly greater coverage of the patched defect by neomucosa in the triple patch group (99.4% vs 93.1% P less than 0.005) and significantly more of the smaller defects were completely covered by neomucosa than the larger defects (12 of 15 vs 0 of 5, P less than 0.05) at 8 weeks. The final area of the defect was 27.5 and 32.8% of the initial patched area respectively for the single and triple patches. Microscopically there was no difference in villous height or crypt depth, but crypt density was significantly greater in the triple group (207 +/- 11 vs 186 +/- 17 crypts/mm, P less than 0.05). In vitro glucose uptake and disaccharidase activity were similar in both groups. Patching multiple small intestinal defects results in more rapid neomucosal growth than a single large defect of the same surface area. This might be due to a greater circumference exposed to surrounding normal mucosa with a resultant increase in crypt density. Since function and villous development of the neomucosa are similar, multiple patches should result in a greater increase in absorptive capacity.
?ds

Set	Items	Description
S1	497	AU="SCOTT M" OR AU="SCOTT M P" OR AU="SCOTT MP"
S2	6	S1 AND PATCHED
S3	4	RD (unique items)
S4	388	PATCHED
S5	157	S4 AND (HUMAN OR MOUSE OR MOSQUITO OR BUTTERFLY OR BEETLE)
S6	14	S5 AND (GENE? OR CLONE? OR DNA?)

```
S7          11   RD (unique items)
S8           4   DROSPHILA AND (GAL OR GALACTOSIDASE)
S9         502   DROSOPHILA AND (GAL OR GALACTOSIDASE)
S10        210   S9 AND EMBRYO?
S11          0   S10A ND DEVELOP?
S12        125   S10 AND DEVELOP?
S13        181   DROSOPHILA(10N)(GAL OR GALACTOSIDASE)
S14         40   S13 AND DEVELOP? AND EMBRYO?
S15         36   RD (unique items)
S16          4   PATCHED AND (GAL OR GALACTOSIDASE)
?s s4 and review?

             388  S4
          391802  REVIEW?
     S17      10  S4 AND REVIEW?
?rd

...completed examining records
     S18       8  RD (unique items)
?t s18/6/1-8


 18/6/1     (Item 1 from file: 154)
09145457   95075457
  Distinct pathways for autocrine and paracrine Wingless signalling in
Drosophila embryos.


 18/6/2     (Item 2 from file: 154)
07988474   92126474
  Blowout of carotid venous patch angioplasty.


 18/6/3     (Item 3 from file: 154)
07961663   92099663
  Ventricular septal defect with tricuspid pouch with and without
transposition. Anatomic and surgical considerations.


 18/6/4     (Item 4 from file: 154)
07944050   92082050
  A review of carotid endarterectomy at a large teaching hospital.


 18/6/5     (Item 5 from file: 154)
07660744   91179744
  Perilymph fistulas: the House Ear Clinic experience.


 18/6/6     (Item 1 from file: 55)
11821956     BIOSIS Number: 98421956
  A consideration of epistemology in systematic biology, with special
reference to species
  Print Number: Biological Abstracts Vol. 100 Iss. 007 Ref. 099548


 18/6/7     (Item 2 from file: 55)
7236094     BIOSIS Number: 38016615
```

GENES THAT CONTROL PATTERN FORMATION DURING DEVELOPMENT

18/6/8 (Item 3 from file: 55)
5862168 BIOSIS Number: 83124475
A CASE OF CONTACT DERMATITIS DUE TO BETAMETHASONE VALERATE WITH A REVIEW OF JAPANESE ARTICLES
?t s18/7/7

18/7/7 (Item 2 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

7236094 BIOSIS Number: 38016615
GENES THAT CONTROL PATTERN FORMATION DURING DEVELOPMENT
SCOTT M P; HAYASHI S; WINSLOW G M; HOOPER J E; SONODA S
DEP. MOL. CELL. DEV. BIOL., UNIV. COLO., BOULDER, COLO. 80309, USA.
CAPECCHI, M. R. (ED.). CURRENT COMMUNICATIONS IN MOLECULAR BIOLOGY: MOLECULAR GENETICS OF EARLY DROSOPHILA AND MOUSE DEVELOPMENT; MEETING, COLD SPRING HARBOR, NEW YORK, USA, APRIL 20-23, 1989. XIII+141P. COLD SPRING HARBOR LABORATORY PRESS: COLD SPRING HARBOR, NEW YORK, USA. ILLUS. PAPER.
ISBN 0-87969-339-8. 0 (0). 1989. 7-10. CODEN: 28365
Language: ENGLISH
?s patched and mammal?

388 PATCHED
137749 MAMMAL?
S19 13 PATCHED AND MAMMAL?
?rd

...completed examining records
S20 12 RD (unique items)
?t s20/6/1-12

20/6/1 (Item 1 from file: 154)
09506686 96028286
Morphogenetic signalling. Responses to hedgehog.

20/6/2 (Item 2 from file: 154)
09394383 95324383
Subcellular localization of the segment polarity protein patched suggests an interaction with the wingless reception complex in Drosophila embryos.

20/6/3 (Item 3 from file: 154)
09081589 95011589
Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments.

20/6/4 (Item 4 from file: 154)
08848979 94163979
A central role for epidermal segment border cells in the induction of muscle patterning in the Drosophila embryo.

20/6/5 (Item 5 from file: 154)
08801069 94116069
Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs.

20/6/6 (Item 6 from file: 154)
08668871 93378871
Contrasting distributions of patched and hedgehog proteins in the Drosophila embryo.

20/6/7 (Item 7 from file: 154)
08460175 93170175
The consequences of ubiquitous expression of the wingless gene in the Drosophila embryo.

20/6/8 (Item 8 from file: 154)
07151658 90058658
The Drosophila patched gene encodes a putative membrane protein required for segmental patterning.

20/6/9 (Item 9 from file: 154)
07108164 90015164
A protein with several possible membrane-spanning domains encoded by the Drosophila segment polarity gene patched.

20/6/10 (Item 10 from file: 154)
07004625 89306625
The role of segment polarity genes during Drosophila neurogenesis.

20/6/11 (Item 11 from file: 154)
06729922 89031922
Patch clamp analysis of Na channel gating in mammalian myocardium: reconstruction of double pulse inactivation and voltage dependence of Na currents.

20/6/12 (Item 12 from file: 154)
06712381 89014381
Patch clamp analysis of chemically activated and modulated ionic channels in isolated mammalian cardiomyocytes.
?

08/319,745

```
=> s patched
L1          1183 PATCHED
=> s patched(5a)(mouse or mammalian or butterfly or beetle)
            1183 PATCHED
           22421 MOUSE
           14401 MAMMALIAN
            7917 BUTTERFLY
            2388 BEETLE
L2             1 PATCHED(5A)(MOUSE OR MAMMALIAN OR BUTTERFLY OR BEETLE)
=> d l2 cit ab
```

1. 4,556,560, Dec. 3, 1985, Methods for the treatment and prophylaxis of diaper rash and diaper dermatitis; Kent W. Buckingham, 424/641; 15/206; 514/494, 502, 865; 604/360 [IMAGE AVAILABLE]

US PAT NO: 4,556,560 [IMAGE AVAILABLE] L2: 1 of 1

ABSTRACT:
Methods for the treatment and prevention of diaper rash and diaper dermatitis caused by the prolonged contact of human skin with body waste are disclosed. The methods of the present invention employ the topical application of a minimum inhibitory concentration of a pharmaceutically-acceptable lipase-inhibiting agent to the area in need of such treatment, or the area where prevention is desired. The lipase-inhibiting agent is preferably a water-soluble metallic salt, such as ZnCl.sub.2, and is preferably applied in combination with a barrier-like vehicle. The effectiveness of these methods is surprising in light of the present confusion and controversy surrounding the actual causes of diaper rash, and the heretofore unrecognized role of lipase as a factor in the cause of diaper rash and diaper dermatitis.

```
=> s patched(5a)(gene# or clon? or DNA)
            1183 PATCHED
            9994 GENE#
           11736 CLON?
           13758 DNA
L3             1 PATCHED(5A)(GENE# OR CLON? OR DNA)
=> d l3 cit ab
```

1. 5,066,596, Nov. 19, 1991, Bacterial strains harboring cloned genes controlling Vibrio cholerae O-antigen biosynthesis; Paul A. Manning, et al., 435/252.33; 424/200.1, 235.1, 242.1, 257.1, 258.1, 261.1; 435/69.1, 69.3, 91.41, 172.1, 172.3, 320.1, 848; 536/24.1; 935/6, 9, 22, 26, 60, 73 [IMAGE AVAILABLE]

US PAT NO: 5,066,596 [IMAGE AVAILABLE] L3: 1 of 1

ABSTRACT:
The invention relates to a fragment of DNA containing genes encoding the synthesis of the O-antigen of Vibrio cholerae serotypes Inaba or Ogawa and being at least 16 kb in length. The invention further related to a cosmid comprising a cloned DNA fragment containing genes encoding the synthesis of O-antigen of Vibrio cholerae serotypes Inaba or Ogawa and to a strain of E.coli that includes the fragment.

```
=> s patched(5a)drosophila
            1183 PATCHED
             974 DROSOPHILA
L4             0 PATCHED(5A)DROSOPHILA
=> s patched and drosophila
```

```
          1183 PATCHED
           974 DROSOPHILA
L5           5 PATCHED AND DROSOPHILA
=> d l5 1-5

1.  5,449,755, Sep. 12, 1995, Human cyclin E; James M. Roberts, et al.,
530/350, 387.1; 536/23.5 [IMAGE AVAILABLE]

2.  5,350,671, Sep. 27, 1994, HCV immunoassays employing C domain
antigens; Michael Houghton, et al., 435/5, 6, 975; 436/512, 518; 530/300,
326, 327, 328, 812, 826; 930/220, 223 [IMAGE AVAILABLE]

3.  5,324,659, Jun. 28, 1994, Yeast mutants useful for indentifying
immunosupressants; Stephen A. Parent, et al., 435/255.2, 942 [IMAGE
AVAILABLE]

4.  5,198,346, Mar. 30, 1993, Generation and selection of novel
DNA-binding proteins and polypeptides; Robert C. Ladner, et al.,
435/69.1, 172.3, 252.3, 320.1 [IMAGE AVAILABLE]

5.  5,096,815, Mar. 17, 1992, Generation and selection of novel
DNA-binding proteins and polypeptides; Robert C. Ladner, et al.,
435/69.1, 172.3, 252.3, 320.1 [IMAGE AVAILABLE]
=> s patched and (gene# or clon? or DNA)
          1183 PATCHED
          9994 GENE#
         11736 CLON?
         13758 DNA
L6           61 PATCHED AND (GENE# OR CLON? OR DNA)
=> s l6 and sequence#
        271178 SEQUENCE#
L7           50 L6 AND SEQUENCE#
=> s l7 and drosophila
           974 DROSOPHILA
L8           5 L7 AND DROSOPHILA
=> d l8 1-5

1.  5,449,755, Sep. 12, 1995, Human cyclin E; James M. Roberts, et al.,
530/350, 387.1; 536/23.5 [IMAGE AVAILABLE]

2.  5,350,671, Sep. 27, 1994, HCV immunoassays employing C domain
antigens; Michael Houghton, et al., 435/5, 6, 975; 436/512, 518; 530/300,
326, 327, 328, 812, 826; 930/220, 223 [IMAGE AVAILABLE]

3.  5,324,659, Jun. 28, 1994, Yeast mutants useful for indentifying
immunosupressants; Stephen A. Parent, et al., 435/255.2, 942 [IMAGE
AVAILABLE]

4.  5,198,346, Mar. 30, 1993, Generation and selection of novel
**DNA**-binding proteins and polypeptides; Robert C. Ladner, et al.,
435/69.1, 172.3, 252.3, 320.1 [IMAGE AVAILABLE]

5.  5,096,815, Mar. 17, 1992, Generation and selection of novel
**DNA**-binding proteins and polypeptides; Robert C. Ladner, et al.,
435/69.1, 172.3, 252.3, 320.1 [IMAGE AVAILABLE]
=> s l7 and (development or embryo?)
        194785 DEVELOPMENT
          4636 EMBRYO?
```

L9 31 L7 AND (DEVELOPMENT OR EMBRYO?)
=> d l9 1-10

1. 5,470,971, Nov. 28, 1995, Stress-induced proteins, **genes** coding therefor, transformed cells of organisms, methods and applications; Keiji Kondo, et al., 536/23.7; 435/69.1, 172.3, 252.3, 254.2, 254.21, 320.1; 536/24.1 [IMAGE AVAILABLE]

2. 5,468,485, Nov. 21, 1995, Avirulent microbes and uses therefor; Roy Curtiss, III, 424/184.1, 93.1, 93.2, 200.1; 435/69.1, 71.1, 172.1, 252.3, 252.33, 252.8 [IMAGE AVAILABLE]

3. 5,449,755, Sep. 12, 1995, Human cyclin E; James M. Roberts, et al., 530/350, 387.1; 536/23.5 [IMAGE AVAILABLE]

4. 5,427,785, Jun. 27, 1995, Rhizosheric bacteria; Clive W. Ronson, et al., 424/93.2; 47/57.6, 58; 71/7; 435/172.3, 252.2, 878; 935/64 [IMAGE AVAILABLE]

5. 5,408,037, Apr. 18, 1995, Methods for detecting glucagon antagonists; Robert A. Smith, et al., 530/308 [IMAGE AVAILABLE]

6. 5,369,766, Nov. 29, 1994, Object-oriented loader system with support for different load formats; Russell T. Nakano, et al., 395/700; 364/280, DIG.1 [IMAGE AVAILABLE]

7. 5,360,901, Nov. 1, 1994, **Gene** **sequence** encoding Aspergillus niger catalase-R; Randy M. Berka, et al., 536/23.2; 435/69.1, 71.1, 172.3, 192, 254.3, 320.1 [IMAGE AVAILABLE]

8. 5,360,732, Nov. 1, 1994, Production of Aspergillus niger catalase-R; Randy M. Berka, et al., 435/192, 69.1, 71.1, 172.3, 254.3, 320.1; 536/23.2; 935/14, 27, 36, 56, 68 [IMAGE AVAILABLE]

9. 5,350,671, Sep. 27, 1994, HCV immunoassays employing C domain antigens; Michael Houghton, et al., 435/5, 6, 975; 436/512, 518; 530/300, 326, 327, 328, 812, 826; 930/220, 223 [IMAGE AVAILABLE]

10. 5,349,127, Sep. 20, 1994, Expression of herbicide metabolizing cytochromes P450; Caroline Dean, et al., 800/205; 435/172.3, 320.1; 800/250, 255, DIG.71; 935/64, 67 [IMAGE AVAILABLE]
=> d l9 1 ab

US PAT NO: 5,470,971 [IMAGE AVAILABLE] L9: 1 of 31

ABSTRACT:
Genes (and portions thereof) which are stress-inducible, e.g., by cold-shock which encode useful proteins. The proteins contribute to confer thermo-tolerance and/or contribute to confer low temperature tolerance to organisms, like eucaryotes or procaryotes. Typical **genes** encoding such proteins and homologous **genes** encoding proteins with equivalent properties are discussed. Nucleotide **sequences** encoding such proteins, recombinant replicable expression vehicles which comprise **DNA** constructs which encode such proteins and competent transformed organisms like eucaryotes are discussed. The production of valuable fermentation products and of biologically active proteins under conditions outside the normal or optimum physiological growth conditions are described.

=> d l9 11-20

11. 5,324,659, Jun. 28, 1994, Yeast mutants useful for indentifying immunosupressants; Stephen A. Parent, et al., 435/255.2, 942 [IMAGE AVAILABLE]

12. 5,321,828, Jun. 14, 1994, High speed microcomputer in-circuit emulator; Michael D. Phillips, et al., 395/500; 364/232.3, 264.3, DIG.1 [IMAGE AVAILABLE]

13. 5,268,274, Dec. 7, 1993, Methods and nucleic acid **sequences** for the expression of the cellulose synthase operon; Arie Ben-Bassat, et al., 435/69.1, 101, 194, 252.3, 252.33, 320.1, 823; 536/23.2; 935/9, 14, 29, 40, 60, 72, 73 [IMAGE AVAILABLE]

14. 5,229,112, Jul. 20, 1993, Combatting plant insect pests with plant-colonizing microorganisms containing the toxin **gene** B. thuringiensis as a chromosomal insertion; Mark G. Obukowicz, et al., 424/93.2; 435/252.3, 252.34 [IMAGE AVAILABLE]

15. 5,212,296, May 18, 1993, Expression of herbicide metabolizing cytochromes; Caroline Dean, et al., 536/23.2; 424/93.2, 93.21; 435/320.1; 536/23.7 [IMAGE AVAILABLE]

16. 5,198,346, Mar. 30, 1993, Generation and selection of novel **DNA**-binding proteins and polypeptides; Robert C. Ladner, et al., 435/69.1, 172.3, 252.3, 320.1 [IMAGE AVAILABLE]

17. 5,190,871, Mar. 2, 1993, Use of the site-specific integrating function of phage .phi.C31; Karen L. Cox, et al., 435/172.3, 252.35, 320.1 [IMAGE AVAILABLE]

18. 5,175,272, Dec. 29, 1992, **DNA** **sequences** with increased expression of HBcAg; Richard L. Mallonee, 536/26.3 [IMAGE AVAILABLE]

19. 5,175,094, Dec. 29, 1992, Increased expression of HBcAg; Richard L. Mallonee, 435/69.3 [IMAGE AVAILABLE]

20. 5,173,427, Dec. 22, 1992, Vectors and hosts with increased expression of HBCAG; Richard L. Mallonee, 435/252.33, 240.2, 252.3, 320.1 [IMAGE AVAILABLE]

=> s p4502c11
L10 0 P4502C11
=> s p4502c?
L11 0 P4502C?
=> s p450?
L12 114 P450?
=> s l12 and hypertension
7510 HYPERTENSION
L13 9 L12 AND HYPERTENSION
=> d l13 1-9

1. 5,294,725, Mar. 15, 1994, Scopularin; Donald R. Kirsch, et al., 549/417 [IMAGE AVAILABLE]

2. 5,288,721, Feb. 22, 1994, Substituted epoxyalkyl xanthines; J. Peter Klein, et al., 514/263; 544/267 [IMAGE AVAILABLE]

08/319,745

```
?b 154 55

       18jan96 08:34:27 User208654 Session D59.1
            $0.12    0.004 Hrs File1
     $0.12  Estimated cost File1
     $0.12  Estimated cost this search
     $0.12  Estimated total session cost   0.004 Hrs.

SYSTEM:OS  - DIALOG OneSearch
  File 154:MEDLINE(R)  1985-1996/Jan W4
         (c) format only 1996 Knight-Ridder Info
  File  55:BIOSIS PREVIEWS(R)  1985-1995/Jan W1
         (c) 1996 BIOSIS

      Set  Items  Description
      ---  -----  -----------
?s pcr(5n)clon?

           43619  PCR
          253749  CLON?
      S1    2678  PCR(5N)CLON?
?s pcr(3n)clon?

           43619  PCR
          253749  CLON?
      S2    2010  PCR(3N)CLON?
?s s2 and review?

            2010  S2
          391802  REVIEW?
      S3       8  S2 AND REVIEW?
?rd

...completed examining records
      S4       6  RD (unique items)
?t s4/6/1-6


 4/6/1     (Item 1 from file: 154)
09461990   95391990
  Clonal  relationship  between  lymphocytic predominance Hodgkin's disease
and concurrent or subsequent large-cell lymphoma of B lineage.


 4/6/2     (Item 2 from file: 154)
08946062   94261062
  Isolated  bone  relapse  during  hematologic remission in childhood acute
lymphoblastic  leukemia:  report  of a metatarsal relapse and review of the
literature.


 4/6/3     (Item 3 from file: 154)
08336712   93046712
  Protein engineering of antibodies.


 4/6/4     (Item 4 from file: 154)
08135143   92273143
```

In vitro antibodies: strategies for production and application.

4/6/5 (Item 1 from file: 55)
11313321 BIOSIS Number: 97513321
The role of molecular genetics in field studies on malaria parasites
Print Number: Biological Abstracts Vol. 098 Iss. 011 Ref. 149060

4/6/6 (Item 2 from file: 55)
10804539 BIOSIS Number: 97004539
A family of serine protease genes expressed in adult buffalo fly (Haematobia irritans exigua)
Print Number: Biological Abstracts Vol. 097 Iss. 001 Ref. 004116
?t s4/7/3-6

4/7/3 (Item 3 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08336712 93046712
Protein engineering of antibodies.
Sandhu JS
Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.
Crit Rev Biotechnol (UNITED STATES) 1992, 12 (5-6) p437-62, ISSN 0738-8551 Journal Code: CRB
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
This article reviews the technical advances in antibody engineering and the clinical applications of these molecules. Recombinant DNA technology facilitates the construction and expression of engineered antibodies. These novel molecules are designed to meet specific applications. Although genomic and cDNA cloning have been used widely in the past to isolate the relevant antibody V domains, at present, the PCR-based cloning is the preferred system. Bacterial and mammalian expression systems are used commonly for the production of antibodies, antibody fragments, and antibody fusion proteins. A range of chimeric antibodies with murine V domains joined to C regions from human and other species have been produced and found to exhibit the expected binding characteristics and effector functions. Humanized antibodies have been developed to minimize the HAMA response, and bifunctional immunoglobulins are being used in tumor therapy and diagnosis. Single chain antibodies and fusion proteins with antibody specificities jointed to nonimmunoglobulin sequences provide a source of antibody-like molecules with novel properties. The potential applications of minimal recognition units and antigenized antibodies are described. Combinatorial libraries produced in bacteriophage present an alternative to hybridomas for the production of antibodies with the desired antigen binding specificities. Future developments in this field are discussed also. (182 Refs.)

4/7/4 (Item 4 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08135143 92273143

In vitro antibodies: strategies for production and application.
Morrison SL
Department of Microbiology and Molecular Genetics, University of California, Los Angeles 90024-1489.
Annu Rev Immunol (UNITED STATES) 1992, 10 p239-65, ISSN 0732-0582 Journal Code: ALO
Contract/Grant No.: CA 16858, CA, NCI; AI29470, AI, NIAID
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, ACADEMIC
The approaches to the production of antibodies (Ab) using the techniques of genetic engineering and expression are reviewed. Genetic engineering facilitates the production of proteins tailormade for an intended use. Bacterial and mammalian expression systems are commonly used for the production of Ab and Ab-like molecules. While genomic or cDNA cloning can be used to obtain the relevant variable regions, PCR-based cloning approaches facilitate the acquisition of additional binding specificities. Large numbers of different chimeric Abs with murine variable regions joined to constant regions from human and other species have been expressed and found to exhibit the expected binding specificities and effector functions. These molecules have been used to study the structural basis of effector functions such as complement activation and Fc receptor binding, and potentially they may be used as therapeutic agents. Carbohydrate has been shown to influence both variable and constant region function. Single-chain Abs and fusion proteins with Ab binding specificities joined to nonimmunoglobulin sequences provide a source of Ab-like molecules with novel properties, and genetically engineered Ab-like molecules provide a source of useful antigens. Combinatorial libraries produced in bacteriophage present an alternative to hybridomas for the production of Abs with desired combining specificities. Issues of the immunogenicity of the recombinant molecules are addressed. (142 Refs.)

4/7/5 (Item 1 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

11313321 BIOSIS Number: 97513321
The role of molecular genetics in field studies on malaria parasites
Walliker D
Div. Biol. Sci., Univ. Edinburgh, West Mains Rd., Edinburgh EH9 3JN, UK
International Journal for Parasitology 24 (6). 1994. 799-808.
Full Journal Title: International Journal for Parasitology
ISSN: 0020-7519
Language: ENGLISH
Print Number: Biological Abstracts Vol. 098 Iss. 011 Ref. 149060
Molecular genetics is having an important impact on the study of genes in natural populations of malaria parasites. The polymerase chain reaction (PCR) is proving particularly valuable for identifying genes in parasites taken directly from their hosts, without the need to establish them in culture. This is leading to novel methods of diagnosis, for example of drug-resistant parasites. Molecular techniques are also greatly assisting understanding of the genetic structure of parasite populations. This is relevant to the current debate on whether Plasmodium falciparum has a clonal or randomly interbreeding structure. Many patients are infected with mixtures of genetically distinct clones. PCR is being used to examine the genotypes of individual oocysts in the mosquito vector. In wild-caught mosquitoes in areas highly endemic for P. falciparum, a large proportion of oocysts are heterozygous, showing that cross-mating occurs frequently

between clones during mosquito feeds. In areas of lower endemicity, there is evidence of less frequent crossing.

4/7/6 (Item 2 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

10804539 BIOSIS Number: 97004539
A family of serine protease genes expressed in adult buffalo fly (Haematobia irritans exigua)
Elvin C M; Whan V; Riddles P W
CSIRO, Div. Tropical Animal Production, Lond Pocket Lab., Private Bag No 3 PO, Indooroopilly 4068, QLD, AUL
Molecular & General Genetics 240 (1). 1993. 132-139.
Full Journal Title: Molecular & General Genetics
ISSN: 0026-8925
Language: ENGLISH
Print Number: Biological Abstracts Vol. 097 Iss. 001 Ref. 004116
Gene fragments encoding serine proteases expressed in adult buffalo fly (Haematobia irritans exigua) were amplified from cDNA using generic oligonucleotide PCR primers, based on conserved residues surrounding the active-site His and Ser amino acids found in all serine proteases. The PCR product consisted of a broad band extending from about 450 bp to 520 bp, which suggested that the PCR product actually consisted of numerous DNA fragments of slightly variable sizes. Seventeen independent clones of these fragments, each with an insert of approximately 480 bp, were digested with HaeIII. Comparison of restriction fragment patterns indicated that 13 of these clones harboured different PCR products. This was confirmed by DNA sequence analysis of 9 clones. Each of the sequenced clones contained all open reading frame which included structurally conserved regions characteristic of the serine protease superfamily. This study reveals the expression of a large and highly variable repertoire of serine proteases in adult buffalo fly. Importantly, these data also demonstrate the utility of such an approach in obtaining DNA probes for use in further investigations of gene family organization and expression, as well as providing recombinant antigens in the form of fusion proteins which may be used as candidates for vaccine production.
?ds

Set	Items	Description
S1	2678	PCR(5N)CLON?
S2	2010	PCR(3N)CLON?
S3	8	S2 AND REVIEW?
S4	6	RD (unique items)

?s s1 and review?

	2678	S1
	391802	REVIEW?
S5	12	S1 AND REVIEW?

?s s5 not s3

	12	S5
	8	S3
S6	4	S5 NOT S3

?rd

...completed examining records
S7 3 RD (unique items)
?t s7/6/1-3

7/6/1 (Item 1 from file: 154)
09095781 95025781
[Low-grade MALT-type non-Hodgkin lymphoma of the stomach with local recurrence 14 years following resection. Demonstration of clonal identity using polymerase-chain-reaction (PCR)]
Niedrig malignes Non-Hodgkin-Lymphom vom MALT-Typ im Magen mit Lokalrezidiv 14 Jahre nach Resektion. Demonstration der klonalen Identitat mittels Polymerase-Kettenreaktion (PCR).

7/6/2 (Item 2 from file: 154)
09081894 95011894
Utility of microsporidian rRNA in diagnosis and phylogeny: a review.

7/6/3 (Item 3 from file: 154)
08677862 93387862
Gene transfer and cardiovascular disorders.
?t s7/7/3

7/7/3 (Item 3 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08677862 93387862
Gene transfer and cardiovascular disorders.
French BA
Department of Medicine, Baylor College of Medicine, Houston, Texas.
Herz (GERMANY) Aug 1993, 18 (4) p222-9, ISSN 0340-9937
Journal Code: F88
Contract/Grant No.: P50-HL42267-01, HL, NHLBI
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
Within the past four years, basic recombinant techniques (such as molecular cloning, sequencing, site-directed mutagenesis, PCR, and transfection) have been combined to yield a "second generation" of recombinant DNA technology with experimental potential which could barely have been envisioned only a decade ago. This review will focus upon the genesis and cardiovascular application of two recent developments in gene transfer technology: gene targeting by homologous recombination and direct in vivo gene transfer. Gene targeting evolved from transgenic mouse technology but is distinguished by its ability to precisely disrupt or "knock-out" specific genes in the murine genome. This not only provides decisive answers to functional questions, but also produces accurate models of human genetic disorders. In vivo gene transfer provides for the direct introduction of genetic information into living tissues. In vivo gene transfer not only facilitates basic research by providing a simple and direct way to analyze gene structure and function in intact animals, but may also find direct clinical application in the treatment of genetic and acquired disorders such as familial hypercholesterolemia and restenosis. (34 Refs.)
?s pcr-based cloning

```
      S8        0  PCR-BASED CLONING
?s pcr(w)based(w)cloning

           43619  PCR
          378332  BASED
           98200  CLONING
      S9      57  PCR(W)BASED(W)CLONING
?rd

...examined 50 records  (50)
...completed examining records
     S10      33  RD (unique items)
?s s10 and review?

              33  S10
          391802  REVIEW?
     S11       2  S10 AND REVIEW?
?t s11/6/1-2


 11/6/1     (Item 1 from file: 154)
08336712   93046712
  Protein engineering of antibodies.


 11/6/2     (Item 2 from file: 154)
08135143   92273143
  In vitro antibodies: strategies for production and application.
?t s10/6/1-33


 10/6/1     (Item 1 from file: 154)
09440942   95370942
  Cloning  of  rat  interleukin-3  receptor  beta-subunit  from  cultured
microglia and its mRNA expression in vivo.


 10/6/2     (Item 2 from file: 154)
09437008   95367008
  Molecular  characterization  of  the  murine  syk protein tyrosine kinase
cDNA, transcripts and protein.


 10/6/3     (Item 3 from file: 154)
09343326   95273326
  Induced  oleoresin  biosynthesis  in  grand fir as a defense against bark
beetles.


 10/6/4     (Item 4 from file: 154)
09291530   95221530
  Cloning  and  subcellular  localization  of  novel  rab  proteins reveals
polarized and cell type-specific expression.


 10/6/5     (Item 5 from file: 154)
09274052   95204052
```

Receptor tyrosine kinases expressed in metastatic colon cancer.

10/6/6 (Item 6 from file: 154)
09245019 95175019
Expression of cyclic nucleotide-gated cation channels in non-sensory tissues and cells.

10/6/7 (Item 7 from file: 154)
09214459 95144459
PCR-based cloning, sequencing, and exon mapping of lymphocyte-derived neuroendocrine peptides.

10/6/8 (Item 8 from file: 154)
09210431 95140431
Enhanced expression of multiple protein tyrosine phosphatases in the regenerating mouse liver: isolation of PTP-RL10, a novel cytoplasmic-type phosphatase with sequence homology to cytoskeletal protein 4.1.

10/6/9 (Item 9 from file: 154)
09207784 95137784
HLA class I allele (HLA-A2) expression defect associated with a mutation in its enhancer B inverted CAT box in two families.

10/6/10 (Item 10 from file: 154)
09115227 95045227
The two nonallelic insulin-like growth factor-I genes in Xenopus laevis are differentially regulated during development.

10/6/11 (Item 11 from file: 154)
09064984 94379984
PCR-based cloning of the full-length Neurospora eukaryotic initiation factor 5A cDNA: polyhistidine-tagging and overexpression for protein affinity binding.

10/6/12 (Item 12 from file: 154)
08953853 94268853
PISSLRE, a human novel CDC2-related protein kinase.

10/6/13 (Item 13 from file: 154)
08905079 94220079
Isolation of three novel POU-domain containing cDNA clones from lactating mouse mammary gland.

10/6/14 (Item 14 from file: 154)
08888677 94203677
Expression of two novel eph-related receptor protein tyrosine kinases in mammary gland development and carcinogenesis [published erratum appears in Oncogene 1994 Aug;9(8):2431]

10/6/15 (Item 15 from file: 154)
08866267 94181267
Molecular cloning of a novel non-receptor tyrosine kinase, HYL (hematopoietic consensus tyrosine-lacking kinase).

10/6/16 (Item 16 from file: 154)
08841206 94156206
Characterization of a novel murine testis-specific serine/threonine kinase.

10/6/17 (Item 17 from file: 154)
08782361 94097361
A single amino acid substitution in the H-2Kb molecule generates a defined allogeneic epitope.

10/6/18 (Item 18 from file: 154)
08754961 94069961
The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5' termini.

10/6/19 (Item 19 from file: 154)
08534794 93244794
The organization of the intron-containing human S6 ribosomal protein (rpS6) gene and determination of its location at chromosome 9p21.

10/6/20 (Item 20 from file: 154)
08452680 93162680
The structure of the human intron-containing S8 ribosomal protein gene and determination of its chromosomal location at 1p32-p34.1.

10/6/21 (Item 21 from file: 154)
08373979 93083979
A small plasmid for recombination-based screening.

10/6/22 (Item 22 from file: 154)
08336712 93046712
Protein engineering of antibodies.

10/6/23 (Item 23 from file: 154)
08241086 92379086
Human inter-alpha-trypsin inhibitor: full-length cDNA sequence of the heavy chain H1.

10/6/24 (Item 24 from file: 154)
08155162 92293162
PCR based cloning and sequencing of isogenes encoding the tree pollen major allergen Car b I from Carpinus betulus, hornbeam.

10/6/25 (Item 25 from file: 154)
08135143 92273143
In vitro antibodies: strategies for production and application.

10/6/26 (Item 26 from file: 154)
08072010 92210010
The complexity of the Rab and Rho GTP-binding protein subfamilies revealed by a PCR cloning approach.

10/6/27 (Item 27 from file: 154)
07821705 91340705
Isolation of a cDNA encoding a mammalian multiubiquitinating enzyme (E225K) and overexpression of the functional enzyme in Escherichia coli.

10/6/28 (Item 1 from file: 55)
11965316 BIOSIS Number: 98565316
Protein tyrosine kinase expression during the estrous cycle and carcinogenesis of the mammary gland
Print Number: Biological Abstracts Vol. 101 Iss. 001 Ref. 008121

10/6/29 (Item 2 from file: 55)
11930714 BIOSIS Number: 98530714
Three new putative voltage gated sodium channels (VGSC) from human neuroblastoma (SH-SY5Y) cells: A PCR based cloning strategy
Print Number: Biological Abstracts/RRM Vol. 047 Iss. 012 Ref. 200589

10/6/30 (Item 3 from file: 55)
11470290 BIOSIS Number: 98070290
Identification of new genes expressed in hemopoietic cell lines: Yield of three PCR-based cloning approaches
Print Number: Biological Abstracts/RRM Vol. 047 Iss. 002 Ref. 031877

10/6/31 (Item 4 from file: 55)
11236030 BIOSIS Number: 97436030
PCR-based cloning and sequence analysis of fixL homologue from Frankia species strain CeS15
Print Number: Biological Abstracts/RRM Vol. 046 Iss. 010 Ref. 160794

10/6/32 (Item 5 from file: 55)
10038487 BIOSIS Number: 95038487
MOLECULAR CLONING AND CHARACTERIZATION OF RKLK10 A CDNA ENCODING T KININOGENASE FROM RAT SUBMANDIBULAR GLAND AND KIDNEY

10/6/33 (Item 6 from file: 55)
7883667 BIOSIS Number: 40084667
PCR-BASED CLONING STRATEGY FOR G PROTEIN-COUPLED RECEPTORS
?t s10/7/4 7-9 11 15 17 24 26 27 30-33

10/7/4 (Item 4 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09291530 95221530
Cloning and subcellular localization of novel rab proteins reveals polarized and cell type-specific expression.
Lutcke A; Parton RG; Murphy C; Olkkonen VM; Dupree P; Valencia A; Simons K; Zerial M
European Molecular Biology Laboratory, Heidelberg, FRG.
J Cell Sci (ENGLAND) Dec 1994, 107 (Pt 12) p3437-48, ISSN 0021-9533
Journal Code: HNK
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Small GTPases of the rab subfamily are specific regulators of vesicular transport. The intracellular localization of these proteins has been mostly investigated in cultured cells where they have been found associated with distinct compartments of the exocytic and endocytic pathways. Using a PCR-based cloning approach we have recently identified several novel rab proteins, extending the total number of this family to more than 30 members. Here, we have investigated the mRNA expression in different tissues and the intracellular localization in organ cryosections of two rab proteins, rab18 and rab20. Both northern blot analysis and confocal immunofluorescence microscopy demonstrated that these proteins are expressed in a tissue- and cell type-dependent manner. Despite their presence in non-polarized cells and polarized cells, both proteins are highly expressed on the apical side of kidney tubule epithelial cells. Electron microscopic studies revealed that rab18 and rab20 are located in apical dense tubules, endocytic structures underlying the apical plasma membrane, suggesting that they play a role in apical endocytosis/recycling. In intestinal epithelial cells as well, both proteins were localized apically, but, in addition, rab18 was found associated with the basolateral domain, suggesting that this protein is not restricted to the apical transport machinery of polarized epithelial cells. The results demonstrate that, depending on the epithelial cell type, rab proteins that are also expressed in non-polarized cells may be enriched in one or both surface domains. Together with the observed tissue- and cell type-dependent variation in the expression of the rab proteins, this suggests that the large number of mammalian rab proteins might reflect the specific requirements in the organization of membrane traffic encountered by different cell types.

10/7/7 (Item 7 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09214459 95144459
PCR-based cloning, sequencing, and exon mapping of lymphocyte-derived neuroendocrine peptides.
Maier CC; Blalock JE
Department of Physiology and Biophysics, University of Alabama at Birmingham 35294.
Immunomethods (UNITED STATES) Aug 1994, 5 (1) p3-7, ISSN 1058-6687
Journal Code: B3R
Contract/Grant No.: P01 NS29719, NS, NINDS; DK38024, DK, NIDDK; T3207335
Languages: ENGLISH
Document type: JOURNAL ARTICLE

In this report a procedure for the analysis of mRNA expression in cells of limited availability by the reverse transcriptase-polymerase chain reaction (RT-PCR) method is described. The cells are lysed with Nonidet P-40, and the mRNA in the lysate is used directly as template for the cDNA synthesis reaction. Target cDNA is then amplified by PCR, and the products can be analyzed that same day by agarose gel electrophoresis. The oligonucleotide primers used for amplification are designed to include restriction sites to facilitate cloning for subsequent sequencing. We have demonstrated that luteinizing hormone-releasing hormone mRNA can be amplified from the hypothalamus and thymus of a 7-day rat pup, in which the starting cell number was limited. Furthermore, exon usage by target cDNA in different cell types can be easily determined by amplifying with exon-specific primers. Proopiomelanocortin (POMC) mRNA expressed in the pituitary utilizes all three exons, while a majority of POMC mRNA expressed in lymphocytes lacks exons 1 and 2. Thus, this provides an extremely rapid and sensitive means not only for analyzing mRNA expression but also for differential exon usage.

10/7/8 (Item 8 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09210431 95140431

Enhanced expression of multiple protein tyrosine phosphatases in the regenerating mouse liver: isolation of PTP-RL10, a novel cytoplasmic-type phosphatase with sequence homology to cytoskeletal protein 4.1.

Higashitsuji H; Arii S; Furutani M; Imamura M; Kaneko Y; Takenawa J; Nakayama H; Fujita J

First Department of Surgery, Faculty of Medicine, Kyoto University, Japan.

Oncogene (ENGLAND) Jan 19 1995, 10 (2) p407-14, ISSN 0950-9232
Journal Code: ONC

Languages: ENGLISH

Document type: JOURNAL ARTICLE

To elucidate the role that protein tyrosine phosphatase (PTPs) may play in liver regeneration, PTPs expressed in the mouse liver after partial hepatectomy (PH) were investigated by a PCR-based cloning method. Sequencing of 115 cDNA clones identified 10 different sequences including MPTP (T cell PTP), PTP-1B, PTP-P19, mR-PTP mu, R-PTP alpha, PTP NE-3 (PTP-P1), R-PTP-kappa and the murine homologue of human LAR. The remaining two sequences, PTP-RL9 and PTP-RL10, encoded novel PTPs. PTP-RL10 cDNA contained an open reading frame of 1176 amino acids with no apparent membrane-spanning region. The amino-terminal region had sequence homology to those of human erythrocyte protein 4.1 and ezrin, cytoskeletal proteins. In the regenerating liver, the levels of five PTP gene mRNAs (MPTP, PTP-P19, R-PTP alpha, LAR homologue, and PTP-RL9) increased within 6 h, decreased to the normal level by 24 h, and increased again at 48 to 72 h after PH. The levels of PTP-1B and R-PTP-kappa mRNAs peaked within 6 h, decreased gradually, and returned to the normal level by 168 h after PH. In contrast, the levels of two PTP mRNAs (mR-PTP mu and PTP-RL10) peaked at 48 to 72 h, and returned to the normal level by 168 h after PH. No expression of PTP NE-3 was detected in the liver by Northern blotting. The differential expression of multiple PTPs during the pre-replicative and post-replicative stages of liver regeneration suggests that PTPs are involved in the regulation of growth and differentiation of liver cells.

10/7/9 (Item 9 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09207784 95137784
HLA class I allele (HLA-A2) expression defect associated with a mutation in its enhancer B inverted CAT box in two families.
Balas A; Garcia-Sanchez F; Gomez-Reino F; Vicario JL
Laboratory of Histocompatibility, Regional Transfusion Center, Madrid, Spain.
Hum Immunol (UNITED STATES) Sep 1994, 41 (1) p69-73, ISSN 0198-8859 Journal Code: G9W
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The present study shows a very highly diminished HLA-A2 cell surface expression with mendelian segregation in two nonrelated Spanish families. The associated haplotype included Cblank-B38-DRB1*1301-DQ6 in both families. cDNA sequence analysis in two members, one of each pedigree, revealed the presence of the commonest HLA-A2 allele (A*0201), without repetitive mutations that could indicate inappropriate or inefficient translation. Further, the coamplified 3'-untranslated region sequence was also the same described for HLA-A2. HLA-A transcription frequency by means of cDNA PCR-based cloning experiments and by Northern blotting pointed out a relatively low number of HLA-A2 mRNA molecules compared with the complementary HLA-A allele. 5'-Regulatory region sequences from two low-expressing HLA-A2 nonrelated individuals showed a unique and identical single point mutation at position -101 (T to C), when compared with all MHC class I alleles sequenced so far. Position -101 is located in the inverted CAT box associated with the MHC class I enhancer B. The fact that this is an extremely well-conserved position leads us to postulate that this change may be the only responsible for the defective expression of HLA-A2.

10/7/11 (Item 11 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09064984 94379984
PCR-based cloning of the full-length Neurospora eukaryotic initiation factor 5A cDNA: polyhistidine-tagging and overexpression for protein affinity binding.
Tao Y; Chen KY
Department of Chemistry, Rutgers University, Piscataway, NJ 08855-0939.
Biochem J (ENGLAND) Sep 1 1994, 302 (Pt 2) p517-25, ISSN 0264-6021 Journal Code: 9YO
Contract/Grant No.: R01 CA49695, CA, NCI
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Eukaryotic initiation factor 5A (eIF-5A) is the only cellular protein known to contain a hypusine residue that is formed by transferring the aminobutyl moiety from spermidine to a specific lysine residue, followed by hydroxylation at the aminobutyl group. A simple PCR-based strategy was developed to obtain a full-length cDNA of Neurospora crassa eIF-5A. The strategy consists of (i) the design of a pair of key primers (21-mer) based on the highly conserved eIF-5A cDNA domains known in other species, (ii) PCR amplification of Neurospora cDNA using the two key primers to obtain the core sequence for the design of core primers, and (iii) combined use of the key primers, core primers and the universal primers, T3 and T7, to

amplify the target sequence in a Neurospora cDNA library. The longest cDNA obtained was cloned into pBlueScript phagemid, and sequence analysis indicated that it encodes a polypeptide of 163 amino acid residues with a codon usage preference characteristic of abundant Neurospora genes. The Neurospora polypeptide showed 59% and 67% identity with human and yeast eIF-5A precursor protein respectively. We subcloned the Neurospora eIF-5A cDNA into pQE-30, which introduces six adjacent histidine residues to the N-terminus of the recombinant protein. The resulting plasmid, pQTy21, was overexpressed in Escherichia coli, and the soluble polyhistidine-tagged protein was purified by metal chelation chromatography. We obtained about 60 mg of purified eIF-5A precursor from 1 litre of culture in a single step using a Ni(II)-nitrilotriacetic acid (NTA)-agarose column. The histidine-tagged eIF-5A precursor protein could be recognized by anti-Neurospora crassa 21 kDa protein serum raised against wild-type eIF-5A precursor and could serve as the substrate protein for deoxyhypusine synthase. Using the histidine-tagged recombinant protein and the Ni(II)-NTA-agarose column, we constructed a protein affinity column and demonstrated an affinity binding between eIF-5A precursor and deoxyhypusine synthase in the presence of NAD+. One-step eIF-5A precursor affinity-column chromatography could lead to a 30-fold purification of deoxyhypusine synthase.

10/7/15 (Item 15 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08866267 94181267
Molecular cloning of a novel non-receptor tyrosine kinase, HYL (hematopoietic consensus tyrosine-lacking kinase).
Sakano S; Iwama A; Inazawa J; Ariyama T; Ohno M; Suda T
Department of Cell Differentiation, Kumamoto University School of Medicine, Japan.
Oncogene (ENGLAND) Apr 1994, 9 (4) p1155-61, ISSN 0950-9232 Journal Code: ONC
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We identified a novel non-receptor tyrosine kinase from a human megakaryoblastic cell line, UT-7, by means of a PCR-based cloning method. The HYL gene contained a SH2 and SH3 domain and a tyrosine kinase catalytic domain. The deduced amino acid sequence of the protein encoded by this gene was most homologous to CSK (c-src kinase). This gene and CSK shared some unique structural properties such as the absence of a myristylation signal and phosphorylation sites of tyrosine residues corresponding to tyrosines 416 and 527 of chicken p60c-src. Unlike CSK, the SH3 domain of HYL was unique since the ALYDY motif was absent. Northern blot analysis revealed a 2.2 kb transcript in various myeloid cell lines but not in adult tissues except for the brain and the lung, whereas CSK mRNA was ubiquitously expressed. The expression of HYL was upregulated when these myeloid cells were differentiated by induction with phorbol myristate acetate. We named this gene, hematopoietic consensus tyrosine-lacking kinase, HYL. The HYL gene was assigned to chromosome 19 at band p13. It is suggested that HYL plays a significant role in the signal transduction of hematopoietic cells.

10/7/17 (Item 17 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08782361 94097361
A single amino acid substitution in the H-2Kb molecule generates a defined allogeneic epitope.
Kesari KV; Van Bleek G; Nathenson SG; Geliebter J
Howard Hughes Medical Institute, Rockefeller University, New York, NY 10021.
Mol Immunol (ENGLAND) Dec 1993, 30 (18) p1671-7, ISSN 0161-5890 Journal Code: NG1
Contract/Grant No.: AI07289, AI, NIAID; AI10702, AI, NIAID
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Using Mitomycin C mutagenesis and negative and positive selection with monoclonal antibodies specific for H-2Kb and H-2Kbm10, respectively, a mutant cell line clone, Mitc-182, was isolated. Direct sequencing of uncloned cDNA as well as PCR based cloning and sequencing of the H-2Kb182 transcript from this mutant revealed a single G-->T transversion resulting in the substitution of Trp167 by cysteine. Serologically, the mutant Kb182 and Kbm10 are almost identical as each has lost at least five Kb specific mAb epitopes and gained several new epitopes. Interestingly, the mutant cell line, Mitc-182, is efficiently recognized by alloreactive CTLs raised in reciprocal combinations, e.g. CB6 anti Cbm10 and Cbm10 anti CB6, indicating that Kb182 contains both Kb and Kbm10 specific epitopes. The mutation has not affected the ability of Kb182 to present Kb restricted antigenic peptides of Sendai and vesicular stomatitis viruses. In addition to underscoring the importance of amino acid residue 167 in alloreactivity, these results indicate a positive correlation between the gain of both an mAb epitope and a defined alloreactive CTL epitope.

10/7/24 (Item 24 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08155162 92293162
PCR based cloning and sequencing of isogenes encoding the tree pollen major allergen Car b I from Carpinus betulus, hornbeam.
Larsen JN; Stroman P; Ipsen H
ALK Research, Horsholm, Denmark.
Mol Immunol (ENGLAND) Jun 1992, 29 (6) p703-11, ISSN 0161-5890 Journal Code: NG1
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Cloning of the gene encoding the major allergen, Car b I, from Carpinus betulus (hornbeam) pollen was performed using the Polymerase Chain Reaction (PCR) to specifically amplify the gene of interest using single stranded cDNA as template. Specific primers, deduced from the aminoterminal sequence of the purified protein, were tailored to facilitate direct expression of plasmic clones, and the large fraction of positive clones obtained, revealed the presence of isogenic variation. Three clones were characterized in detail by antibody based assays and nucleotide sequencing. The recombinant allergens were shown by crossed immunoelectrophoresis (CIE) to precipitate with monospecific polyclonal rabbit antibodies raised against purified Bet v I, by crossed radioimmunoelectrophoresis (CRIE) to bind tree pollen allergic patient serum IgE, and by immunoblotting to bind murine monoclonal antibodies, raised against purified Car b I from pollen. Car b I is encoded by a 159-triplets open reading frame. The molecular masses (M(r) = 17272, 17355 and 17217 Da, respectively), the amino acid

composition, and the aminoterminal sequence of the predicted polypeptides agree well with data obtained by analysis of the protein purified from pollen. The deduced amino acid sequences show pronounced homology (73, 75 and 74% identities respectively) to Bet v I, the major allergen from Betula verrucosa (white birch) pollen. Soluble recombinant Car b I, without a fusion partner, was produced in Escherichia coli with an immunochemical reactivity closely resembling that of the native pollen allergen. The tree pollen major allergens therefore constitute an ideal system for the study of allergenic epitopes.

10/7/26 (Item 26 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

08072010 92210010
The complexity of the Rab and Rho GTP-binding protein subfamilies revealed by a PCR cloning approach.
Chavrier P; Simons K; Zerial M
Cell Biology Programme, European Molecular Biology Laboratory, Heidelberg, F.R.G.
Gene (NETHERLANDS) Mar 15 1992, 112 (2) p261-4, ISSN 0378-1119 Journal Code: FOP
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Partial sequences corresponding to eleven novel Rab proteins and one new Rho protein have been isolated using a PCR-based cloning approach. These results confirm that the overall diversity of the Rab and Rho protein subfamilies account for more than thirty different members in mammalian cells.

10/7/27 (Item 27 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

07821705 91340705
Isolation of a cDNA encoding a mammalian multiubiquitinating enzyme (E225K) and overexpression of the functional enzyme in Escherichia coli.
Chen ZJ; Niles EG; Pickart CM
Department of Biochemistry, State University of New York, Buffalo 14214.
J Biol Chem (UNITED STATES) Aug 25 1991, 266 (24) p15698-704, ISSN 0021-9258 Journal Code: HIV
Contract/Grant No.: AI28824, AI, NIAID
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The ubiquitin (Ub)-conjugating enzyme E2(25K) catalyzes the synthesis of multi-Ub chains in which successive Ub units are linked by an isopeptide bond involving the epsilon-amino group of Lys-48 of Ubn, and the COOH-terminal Gly residue of Ubn+1 (Chen, Z., and Pickart, C. M. (1990) J. Biol. Chem., 265, 21835-21842). We now describe the polymerase chain reaction (PCR)-based cloning of an E2(25K)-encoding cDNA from a bovine thymus library, using degenerate oligonucleotide primers based on the sequences of two E2(25K) peptides. The cDNA encodes a 200-residue protein whose sequence bears similarities of 66 and 59%, respectively, to the sequences of the Ub-conjugating enzymes encoded by the UBC1 and UBC4/UBC5 genes of the yeast Saccharomyces cerevisiae. These three yeast E2s play key roles in Ub-dependent proteolysis (Seufert, W., McGrath, J. P., and

Jentsch, S. (1990) EMBO J. 9, 4535-4541). Comparison of the amino acid sequence of E2(25K) with other known E2 sequences strongly suggests that Cys-92, one of two E2(25K) Cys residues, forms the Ub thiol ester adduct that is an intermediate in E2-catalyzed multiubiquitination. The E2(25K)-encoding cDNA was overexpressed in Escherichia coli, and the recombinant E2(25K) protein was purified to electrophoretic homogeneity; enzymatic assays showed that its multiubiquitinating activity was quantitatively identical with that of the native protein. The availability of a cloned cDNA will allow us to assess the physiological role of E2(25K).

10/7/30 (Item 3 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

11470290 BIOSIS Number: 98070290
Identification of new genes expressed in hemopoietic cell lines: Yield of three PCR-based cloning approaches
Furukawa T; Nakamoto B; Papayannopoulou T; Stamatoyannopoulos G
Div. Hematol., Univ. Wash., Seattle, WA, USA
Blood 84 (10 SUPPL. 1). 1994. 414A.
Full Journal Title: Abstracts Submitted to the 36th Annual Meeting of the American Society of Hematology, Nashville, Tennessee, USA, December 2-6, 1994. Blood
ISSN: 0006-4971
Language: ENGLISH
Print Number: Biological Abstracts/RRM Vol. 047 Iss. 002 Ref. 031877

10/7/31 (Item 4 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

11236030 BIOSIS Number: 97436030
PCR-based cloning and sequence analysis of fixL homologue from Frankia species strain CeS15
Murry M A; Stigter J; De Bruijn F J
Dep. Bot. and Plant Pathol., Mich. State Univ., East Lansing, MI 48824, USA
0 (0). 1993. 485.
Full Journal Title: Palacios, R., J. Mora and W. E. Newton (Ed.). Current Plant Science and Biotechnology in Agriculture, Vol. 17. New horizons in nitrogen fixation; 9th International Congress on Nitrogen Fixation, Cancun, Mexico, December 6-12, 1992. xvi+788p. Kluwer Academic Publishers: Dordrecht, Netherlands; Norwell, Massachusetts, USA. ISBN 0-7923-2207-X.
ISSN: *********
Language: ENGLISH
Print Number: Biological Abstracts/RRM Vol. 046 Iss. 010 Ref. 160794

10/7/32 (Item 5 from file: 55)
DIALOG(R)File 55:BIOSIS PREVIEWS(R)

10038487 BIOSIS Number: 95038487
MOLECULAR CLONING AND CHARACTERIZATION OF RKLK10 A CDNA ENCODING T KININOGENASE FROM RAT SUBMANDIBULAR GLAND AND KIDNEY